**Lageparameter**: xD (Modus, Modalwert), xZ(Median, Zentralwert, x̅ (Mittelwert), Quantil

“Lage” der Elemente der GGH / Stichproben bezgl. Messskala, keine Aussage über Daten-Streuung,

**Streuungsparameter**: w (Spannweite), IQR, s² (Varianz), s (Standardabweichung), v (Varianzkoeffizient)

**linksschief (rechtssteil)**: XD > XZ > x̅ , **rechtsschief** (linkssteil): XD < XZ < x̅, metrische Verteilung xZ = x̅

2-dimensionale Häufigkeitsverteilung 🡆 Kreuztabelle

M = Merkmal (A oder B), G = Geschlecht

für klassierte Daten

**ri =**

**hi = ri \* bi**

n = ∑hi = h1 + h2 +h3

20 = 16 + h2 + h3  | | -16

4 = h2 + h3 | - h3

h2 = 4 – h3

x̅ = ∑(xi \* hi)

1,25 =  **\*** 16 + 2h2 +3 h3 |

25 = 16 + 2 \* (4-h3) + h3 | - 16

9 = 8 **– 2h3 + h3** | - 8

1 = **– 2h3 + h3  🡆 h3 = 1**

h2 = 4 – h3  = 4 – 1 = 3

Randverteilung für eindim.. Häufigkeitsverteilung von G

absolute Häufigkeit

folgende Werte sind in d. Tabelle einzutragen

*relat. Spaltenhäufigkeit (z. B 400 von 1000)*

*relat. Zeilenhäufigkeit (z. B 400 von 1200)*

*% unter den Summen sind relative Werte der Zeilen- / Spaltensumme zu den Randverteilungen* (Gesamtsumme)

|  |  |  |  |
| --- | --- | --- | --- |
| M/G | m | w | ∑ (**Randverteilung**) |
| A | **400**  *40%*  *33,33%* | **800**  *80%*  *66,66%* | *1200*  60% v **2000** |
| B | **600**  *60%*  *75%* | **200**  *20%*  *25%* | *800*  40% v **2000** |
| ∑ | *1000*  *50%* v. **2000** | *1000*  *50%* v. 2000 | ***2000***  100% |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Merkmals-  ausprä-  gung  Anzahl der Handys | absolute Häufigk.  Anz. Handy  -Nutzer |  | relative Häufigk.  **hi / n** | abs. Summen-  häufigk.  **hi+hi+1** | rel. Summen-  häufigk.  **fi + fi+1** | arithm.  Mittel  **∑(xi \* hi) / n**  **25 / 20** | Zwischen-  rechnung für Varianz |
| xi | hi | xi \* hi | fi | Hi | Fi | x̅ | (xi – x̅)2 \* hi |
| **1** | **16** | 16 | **0,80** | 16 | **0,8** | **1,25** | 1 |
| **2** | 3 | 6 | 0,15 | 19 | 0,95 | **1,25** | 1,687 |
| **3** | 1 | 3 | 0,05 | **20** | **1,00** | **1,25** | 3,063 |
| ∑ | **n = 20** | **25** | **1,00** | - | - | - | **5,75** |

**Summe xi \* hi SHIFT** 🡆 Taste **„1“** 🡆 **„3: Sum“ 🡆 „2: ∑x**

**Mittelwert x̅ SHIFT** 🡆 Taste **„1“** 🡆 **„4: Var“** 🡆 **„2: x̅“ 🡆 x̅2**

**Mittelwert bei klassierten Daten** wie bei unklassierten Daten nur mit Klassenmitte mi, statt mit xiarithmetische Mittel aus unterer und oberer Klassengrenzem**i (**Klassenmitte: **= \* hi** = absol. Häufigkeit der Klasse

bei klassierten Daten können x̅ , Varianz s², Std.-Abw. s, Varianzko-effizient v wie bei unklassierten Daten im Casio berechnet werden.

Statt xi wird mi (Klassenmitte) in Tabelle eingetragen

Klassenbreite Bsp. Klasse „20 b. u. 35 “: Klassenbreite bi = 35 – 20 = 15

**Varianz s² SHIFT** 🡆 Taste **„1“** 🡆 **„4: Var“** 🡆 **„3: Sigma x“ 🡆 „x²“**

**arithmetische Mittel der Abweichungsquadrate auf Grund der Quadrierung immer ≥ 0**

wichtiger Streuungsparameter, für metrische Merkmale, Ausgangswert für Standardabweichung und Variationskoeffizient

**s² = 1/n \* ( (x1 – x̅)² \* h1 + (x2 – x̅)² \* h2 + (x3 – x̅)² \* h3 )**

**s² = 1/20 \* ( (1 - 1,25)2 \* 16 + (2 – 1,25)² \* 3 + (3 – 1,25)² \* 1 )**

**Varianz s² bei klassierten Daten**

wie bei unklassierten Daten nur mit Klassenmitte mi statt mit xi

**Standardabweichung (durchschn. Abweichung) s SHIFT 🡆 Taste „1“ 🡆 „4: Var“ 🡆 „3: Sigma x“**

**Varianzkoeffizient v**

**v = s / x̅** **Standardabweichung s / Mittelwert x̅**

Relatives Streuungsmaß (Streuungsparameter), dimensionslose Größe, prozentuales Verhältnis der Standardabweichung zum arithmetischen Mittel, zum Vergleich der Streuung zwischen verschiedenen Erhebungen

**Modus x̅D** 🡆 Beobachtungswert mit der größten Häufigkeit 🡆 mehrere Merkmalsausprägungen mit der gleichen maximalen Häufigkeit = mehrere Modalwerte = Beobachtungswert mit Häufigkeit == 1 = kein Modus!) 🡆 Modus xD aus o. g. Tabelle ist 1 🡆 x1 = 1 hat die max. hi = 16

bei klassierten Daten ist Modalwert die Mitte der Klasse mit den größten Häufigkeiten (Modalklasse)

im o. g. Bsp. ist Klasse k2 mit h2 = 20 die Modalklasse und der Modalwert = mi = ½ \* (35+55) = 45

**Spannweite w**

w = xmax - xmin **auch bei klassierten Daten!**

**Quartile** 🡆 Q1, Q2, Q3 🡆 Q2-Quartil = **Median / Zentralwert x̅Z Beobachtungs-Werte müssen geordnet sein!**

wenn xn ungerade **x̅z =** wenn xn gerade: **x̅z =**

**Median xZ:** mind. 50% der Beobachtungswerte liegen links und rechts des Medians

n ungerade xn+1/2 🡆 **x-Wert aus der Zeile nehmen, wo Hi ≤ xn+1/2** Bsp. n = 25, XZ aus Zeile, wo Hi ≤ 13 ((25+1) /2)

n gerade (xn/2 +  xn/2+1) /2, 🡆 **x-Wert aus der Zeile nehmen, wo Hi ≤ xn/2+1** Bsp. n = 26, Xi aus Zeile, wo Hi ≤ 13 ((26 /2) und Hi ≤ 14 (26 /2 + 1) beide xi-Werte addieren und durch 2 dividieren

**bei klassierten Daten**

k = Einfallsklasse 🡆 Klasse mit rel. Summenhäufigkeit Fi ≥ 50%,

**x**k-1 = Untergrenze d. Einfallsklasse, **x**k = Obergrenze d. Einfallsklasse,

Fk-1 = relative Summenhäufigkeit der Klasse unterhalb (vor) der Einfallsklasse

fk = relative Häufigkeit der Einfallsklasse

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| xi | hi | fi | Hi | Fi |
| 10 b. u. 30 | 60 | 30 | 60 | **30** |
| **30** b. u. 40 | 80 | 40 | 140 | 70 |
| 40 b. u. 70 | 60 | 30 | 200 | 100 |
| ∑ | 200 | 100 | - | - |

**Einfallsklasse ist k2 „30 b. u. 40“, da Fi ≥ 50** (in Klasse „10 b. u. 30“ ist Fi < 50

**Hinweis. %-Werte immer als Dezimalwerte schreiben**

**xk-1 = K-Untergrenze, xk = K-Obergrenze**

bei Q1, Q3 wird genauso verfahren.

Die Einfallsklasse muss rel. Summenhäufigkeit

Fi ≥ 25% (bei Q1) bzw. Fi ≥ 75% (bei Q3) haben.

Für den %-Wert im Zähler des Bruchs wird statt

0,5 🡆 0,25 (für Q1) bzw. 0,75 (für Q3) genommen

**Interquartilsabstand IQR = Q3 – Q1**

Streuungsmaß

Differenz zwischen oberem und unterem Quartil

enthält 50% der Verteilung.

**Boxplot**: enthält Min, Q1, Median, Q3, Max  
informiert über Streuungsmaße Spannweite (xmax – xmin), IQR

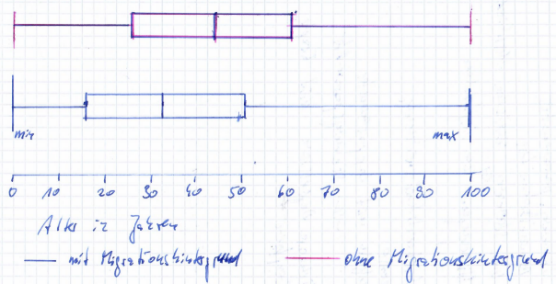
Schiefe und Ausreiße

Achsenbeschriftung und Legend („mit Migrationshintergr.“) nicht vergessen, beides unterhalb das Boxplots“

Achse enthält die Skala zu den x-Werten.

Boxplot enthält die zum Median und zu den Quartilen berechneten x-Werte und Xmax- und Xmin sowie den IQR

**auch bei klassierten Daten**



**Korrelationsanalyse 🡆 Zusammenhangsanalyse (Interdependenzanalyse) zwischen 2 metrischen Merkmalen X, Y**K-Analyse prüft, **ob zwei Variablen X und Y linear zusammenhängen** und **prüft die Stärke des Zusammenhangs**

**analysiert Wechselwirkung der Variablen X, Y untereinander** (wie beeinflussen sich die Variablen gegenseitig?)  
**Korrelationskoeffizient rx,y: Zusammenhangsmaß (Assoziationsmaß) für Quantifizierung der Korrelation-1 ≤ rxy ≤ +1**

**Korrelation 🡆 zahlenmäßiger statistischer Zusammenhang zwischen 2 Merkmalen X und Y**

positive K. ( **r > 0)**: **gleichförmige Entwicklung von X und Y (hoher Wert von X 🡆 auch hoher Wert von Y**)

negative K. (**r < 0)**: **gegenläufige Entwicklung von X und Y (hoher Wert von X 🡆 jedoch niedriger Wert von Y)**

r = -1: extrem starker negativer linearer Zusammenhang 🡆 Punktewolke mit negativer Steigung (l. o. nach r. u.)

r = +1: extrem starker positiver lin. Zusammenhang 🡆 Punktewolke mit positiver Steigung.; r = 0: kein lin. Zusammenhang

**Korr.** **trifft** aber **KEINE Aussage zum kausalen Zusammenhang und zur Kausalitätsrichtung**, jedoch Voraussetzung für Erklärung eines kausalen Zusammenhangs (kausaler Zusammenhang 🡆 ist Ursache-Wirkungs-Beziehung zwischen X und Y 🡆 Veränderung des abhängigen Merkmals Y basiert auf Veränderung von X

**Korrelationskoeffizient rxy: 🡆 Wertebereich: -1 ≤ rxy ≤ +1**

**- dimensionsloses Maß für den Grad des linearen Zusammenhangs** (**Stärke und Richtung des lin. Zusammenhangs**)

**K-Koeffizient r für 2 mind. intervallskalierte Merkmale X und Y , m. positiver Kovarianz cov(x,y) & Standardabweichung s**

**K-Koeffizient r berechnet sich aus dem Quotienten aus Kovarianz cov(x, y): und Standardabweichung s**

**für Kovarianz cov(x, y) im Zähler**

bezieht sich auf die Summe des zuvor zu (xi \* yi) berechneten Produkts (**nicht die Summen von x und y multiplizieren, sondern das in jeder Tabellen-Zeile berechnete Produkt aus (xi \* yi) addieren (hier 450) !**

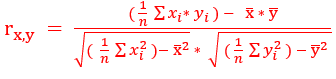
**für Standardabweichung s im Nenner**   
gilt dasselbe. Quadrate für xi, yi summieren (= 74 und 3000) und x̅ ² bzw. y̅² subtrahieren.

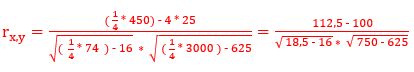
**Bsp:**

**Werte in Tabelle erfassen:**

**xi, yi**  & **xi \* yi**  *&* **xi², yi²** & **Summen zu allen Werten** & **Mittelwert zu xi und yi**  &  **Quadrate zu Mittelwerten**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Filial-Nr  i | | xi  VK-Fläche  Tsd. m² | yi  Umsatz  (Mio €) | xi² | yi² | xi \* yi | |
| 1 | | 3 | 30 | 9 | 900 | 90 | |
| 2 | | 2 | 10 | 4 | 100 | 20 | |
| 3 | | 6 | 40 | 36 | 1600 | 240 | |
| 4 | | 5 | 20 | 25 | 400 | 100 | |
| **n = 4** | **∑** | **16** | **100** | **74** | **3000** | | **450** |
| **x̅ = 4** (**16 : 4**)  **x̅² = 16** | | | **y̅ = 25** (**100 : 4**)  **y̅² = 625** | | | | | |







**Interpretation d. Korrelationskoeffizienten rxy:** nur Orientierungswerte → Entscheidung ist immer problembezogen

**-1 bis -0,8 „starke“ negative Korrelation, -0,8 bis -0,5 „mittlere“ neg. K., -0,5 bis 0 „schwache“ K., 0 keine Korrelation**

**0 bis 0,5 „schwache“ positive K., 0,5 bis 0,8 „mittlere“ K, 0,8 bis 1 „starke“ positive K.**

**Regressionsanalyse 🡆 Abhängigkeitsanalyse (Dependenzanalyse) für gerichteten Zusammenhang**

ist Abhängigkeitsanalyse zwischen 2 Merkmalen X, Y ; Unterscheidung zwischen unabhängigen und abhängigen Merkmalen

R-Analyse basiert auf Regressionsfunktion ŷ = a + b\*x 🡆 R-Funktion ermittelt Abhängigkeitsmaß

**einfache Regressionsanalyse** mit 2 metrischen Größen: Zielgröße Y ist Regressand (abhängiges Merkmal), Einflussgröße X ist Regressor (unabhängiges Merkmal); **Zweck: zum Erstellen eines Vorhersagemodells, Quantifizierung der Stärke des Zusammenhangs**

**Regressionsfunktion (Dependenzanalyse)**

Voraussetzungen: X, Y sind quantitative (metrische) Merkmale, linearer Zusammenhang zwischen X→ Y

**Regressionsfunktion ŷ = a + b \* x (nach Methode der kleinsten Quadrate)**

anhand der **2 Regressionskoeffizienten a, b** **veranschaulicht** eine **Gerade in Punkwolke** des Streu-Diagramms den linearen **Zusammenhang;**

**zu den Regressionskoeffizienten a und b (Kurvenparameter)** soll die Summe der quadratischen Abweichungen der Kurve von den beobachteten Punkten minimal sein

Berechnung des **Regressionskoeffizienten a** Berechnung des **Regressionskoeffizienten b**

Formel im Nenner ist in beiden Gleichungen identisch

Bsp. zum Zusammenhang von VK.-Fläche und Umsatz 🡆 Tabelle mit Hilfsgrößen

**Interpretation der R-Koeffizienten a und b:**

a ist konstante Größe und wird durch keine Einflussgröße beeinflusst, a ist unabhängig von Einflussgröße,

b ist variabler Faktor im veränderlichen Term der R-Rechnung, b ist abhängig von Einflussgröße (unabhängiger Variable) x

**ŷ = a + b \* x ŷ = 5 + 5 \* x**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Filialen-Nr.  i |  | xi  VK-fläche  (1000 qm) | yi  Umsatz  (Mio €) | xi² | yi² | xi \* yi | **ŷ**  = **a** + b \* x  = **5** + **5 \* x** |
| 1 |  | **3** | 30 | 9 | 900 | 90 | **20** |
| 2 |  | **2** | 10 | 4 | 100 | 20 | **15** |
| 3 |  | **6** | 40 | 36 | 1600 | 240 | **35** |
| 4 |  | **5** | 20 | 25 | 400 | 100 | **30** |
| **n = 4** | **∑** | **16** | **100** | **74** | **3.000** | **450** | **100** |

**Varianz d. Residuen**: in Hilfstabelle 1 Spalte mit werten aus **ŷ** **= a + b \* x** 🡆 Quadrate jedes **ŷ** addieren und durch n dividieren; davon Quadrat zum Mittelwert von subtrahieren 🡆 **¼ \* (20² + 15² + 35² + 30²) – 25²**

Varianz der Regressionswerte wird durch die Varianz d. unabhängigen Merkmals bestimmt

**Lineare Korrelation und Regression erfassen nur lineare Zusammenhänge und sind anfällig für Ausreißer**

Im **Streudiagramm**: Punkte (kleine Vierecke) anhand der Koordinaten x, y aus d. Tabelle zeichnen (**Achsbeschriftung**!)

Für Regressionsgerade 1 Punkt zur Y-Achse anhand der Regressionsfunktion berechnen. Dabei einen Wert für X aus Tabelle einsetzen. Für den 1. Punkt der Gerade wird x = 0 eingesetzt 🡆 ŷ dann = R-Koeffizient a, x = 0. Für 2. Punkt: y = zuvor berechneter Punkt (**ŷ**  **= a + b \* x** ). Beide Punkte zu Y und X im Diagramm zeichnen und verbinden 🡆 R-Gerade muss über das gesamtes Diagramm gehen

**Bestimmtheitsmaß R²= (rx,y)²**  🡆 **immer in % angeben!** (Quadrat der Korrelationskoeffizienten r in %) **immer ≤ 0 ≤ 1**

**0% ≤ R² ≤ 100% ( bzw. 0 ≤ R² ≤ 1)** aus o. g. Bsp: rx,y = 0,707 🡆 R²= (rx,y)² = 0,707² = 0,50 = 50%  
Gütemaß der lin. Regression (Erklärungskraft d. Modells: wie gut entspricht das Modell der Realität R² = 100% = perfektes Modell); wie gut erklärt die unabhängige Variable (Regressor) X, die Varianz der abhängigen Variable Y (Regressand) = der durch Varianz des unabhängigen Merkmals X erklärbare Anteil der Varianz d. abhängigen Merkmals Y (Residualvarianz) 🡆 Varianz d. abhängigen Merkmals ist in erklärbare und nicht erklärbare Varianz zerlegbar  
***Antwort um Bsp. 50% der Varianz zu den Umsätzen lassen sich durch Varianz der Verkaufsflächen erklären. Die übrigen 50% der Varianz in den Umsätzen sind durch andere Einflussgrößen erklärbar.***

R² im Diagramm: je mehr Datenpunkte auf einer Linie, umso höher R². Streuen Datenpunkte o. Zusammenhang, ist R² nahe 0

Umsatzprognose bei Erweiterung der VK-Fläche um 1000 qm in Filiale 2

**ŷ = 5 + 5 \* 2 + 1 = 5 + 15 = 20**

Bei einer Erweiterung der VK-Fläche um 1000 qm in Filiale 2 erhöht sich der Umsatz von 10 Mio auf 20 Mio €.

**Wahrscheinlichkeitsrechnung**

Ω= Ereignisr., gesamte Ergebnismenge, n = Zahl der möglichen Fälle, k = Elemente der Menge = Zahl d. günstigen Fälle

*Aus* ***3 Parteien A, B. C*** *mit* ***jeweils 7 Mitgliedern*** *soll ein* ***6-köpfiger Ausschuss*** *gebildet werden.*

*Wie viele Ausschusszusammensetzungen gibt es insgesamt?*

**ohne Reihenfolge, ohne Wiederholung = KOW = Binomialkoeffizient** Bestimmung der k-Objekte aus Menge n (n aus k)  
alle 3 Parteien haben gesamt 7 \* 3 = 21 Mitglieder 🡆 für 6 Ausschüsse sind die Varianten (Zusammensetzungen) zu ermitteln

Ermittlung über Binomialkoeffizient 𝛀 (Gesamtmenge der Ausschusszusammensetzungen):*Wie viele Ausschusszusammensetzungen, wenn mindestens ein A-Mitglied dabei sein soll?*

Denkweise „verkehrt“! „mindestens ein…“ = Gesamtmenge (21) – „ohne A“ 🡆 Von Ω wird der Anteil der |B| und |C|-Mitglieder (2 x 7 = 14) subtrahiert

**|Ω| - |A| = = 51.261** 🡆 im Casio: 21 🡆  🡆  🡆 6 🡆 – 14 🡆  🡆  🡆 6

*Wie viele Ausschusszusammensetzungen, wenn mindestens ein A-Mitglied und ein B-Mitglied dabei sein sollen?*

**Siebformel für 2 Mengen**  **|𝜴| − ( |𝑨| + |𝑩| - |𝑨 ∩ 𝑩| )** **|𝑨 ∩ 𝑩| 🡆 Schnittmenge von A und B =**

**im Casio:** 21C6 – 2 x 14C6 + 7C6

*Wie viele Ausschusszusammensetzungen, wenn aus jeder der drei Parteien mindestens ein Mitglied dabei sein soll*

**Siebformel für 3 Mengen** |**𝜴| - ( |𝑨| + |𝑩| + |C| - |𝑨 ∩ 𝑩| - |A ∩ C| - |B ∩ C| + |A ∩ B ∩ C| )**

**im Casio:** 21C6 – 3 x 14C6 + 3 \* 7C6

**P(A) = 1–P(A̅) = 1 – P(B1) + P(B2)+P(B3) – P(B1⋂B2) – P(B1⋂B3) – P(B2⋂B3 + P(B1⋂B2⋂B3)** 🡆*Studenten mit versch. Kursen*

*Wie viele „Worte“ können ELEVEN aus dem Wort ELEVEN erzeugt werden?*

**Permutat. m Wiederholung; Casio-Funktion nPr (Taste** **)** 🡆 **bei mehr als 2 B-Typen sind Fakultäten zu dividieren**

**z. B.** ELEVEN = **6 Buchstaben, 4 B-Typen** 🡆 **3 x „E“, 1 x „L“, 1 x „V“, 1 x „N“**  =

*Wie viele dieser „Worte“ enthalten die drei E's direkt hintereinander?*

**Sonderfall:** „E“ soll hintereinander folgen 🡆 **3 E's = 1 Zeichen** 🡆 „**ELVN**“, keine Fakultät im Nenner 🡆 4! = 24

*c) Wie viele dieser „Worte“ beginnen mit E und enden mit N?*

„E“ und „N“ am Beginn / Ende des Wortes (E)**LEVE**(N) werden nicht berücksichtigt 🡆 „**LEVE**“ **🡆 n=** **4** und

**3 B-Typen**: **1 x „L“**, **2 x „E“**, **1 x „V“** 🡆 1! \* 2! \* 1! = ODER **;**

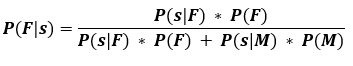
Permutat. m Wiederholung auch zur Anzahl der Kombis für Perlen die auf Kette aufgefädelt werden

**bedingte Wahrscheinlichkeit 🡆** Wahrscheinlichkeit für Ereignis A nach Ereignis B

**„Satz der totalen Wahrscheinlichkeit“:**  Ermittlung P(A) unter **EINER Voraussetzung P(B )**

**P(A) = ∑P(A|B) \* P(B) = P(A|B) \* P(B)** + **P(A|B̅ ) \* P(B̅)** 🡆 Baumdiagr.: 2 Pfade, pro Pfad multiplizieren, Produkte addieren

**Satz von Bayes**: Ermittlung P(A) unter **ZWEI Voraussetzungen**: P(A|B) und P(B|A)



**A ∪ B = Vereinigungsmenge** 🡆 Objekte in **A ODER B** 🡆 **( x ∈ A ) ∨ ( x ∈ B )** 🡆 **Addition**

**A ∩ B = Schnittmenge** 🡆 Objekte in **A UND B** 🡆 **( x ∈ A ) ∨ ( x ∈ B )** 🡆 **Multiplikation**

**A \ B = Differenzmenge** 🡆 Objekte, in **A aber NICHT in B** 🡆 ( **x ∈ A ) ∧ ( x ∉ B )** 🡆 **Subtraktion** A - B

**BC oder B̅ Komplement v. B in Bezug auf A** 🡆 **A OHNE B** 🡆 **{ x ∣ x ∉ B } 🡆 Gesamtmenge ohne Elemente (Objekte) aus B**

🡆 Subtraktion: **Ω = 1 - |B|** (= Komplementärmenge 🡆 Gegenwahrscheinlichkeit)

**mit Reihenfolge, mit Zurücklegen 🡆** VmW 🡆 **|Ω| =** 𝒏𝒌

**mit Reihenfolge, ohne Zurücklegen 🡆** VoW 🡆  **Sonderfall wenn n = k, dann |Ω| = n!**Bruch kürzen, falls n sehr groß 🡆 Taschenrechner kein Ergebnis liefert 100! / (100 -3)! = 100! / 97! = 100 \* 99 \* 98

**ohne Reihenfolge, mit Zurücklegen** 🡆 KmW 🡆

**ohne Reihenfolge, ohne Zurücklegen** 🡆 KoW 🡆 (Binomialkoeffizient); k Elemente aus Menge n

günstige Fälle |E| / mögliche Falle |Ω|; **p(A)= E / Ω**

*Wahrscheinlichk. zur Los-Auswahl v. 5 Angriffsspielern bei 10 Spielern*:

*Wahrscheinlichkeit, viermaliges Werfen* ***eines*** *Würfels verschiedene Zahlen? 🡆 P m. Reihenfolge m. Zurücklegen, n = 6, k = 4*

**1.** |Ω| 🡆 **VmW**= **nK** **= 1296** **2.** |E| 🡆 P = **Vow**  **3.** **p = E / Ω** = 3 / 1296

*Selbiger Algorithmus bei „Wörter der Länge 3 aus 5 Buchstaben. P für Wörter mit nur 2 verschiedene Buchstaben?*

**Baumdiagramm (Wahrscheinlichkeitsgraph)** mögliche Wahrscheinlichkeiten als Linien mit Knoten am Ende, an den Linien die Wahrscheinlichkeiten notieren, bei der ersten Wahrscheinlichkeit ausspalten in neue mögliche W. (neue Linien), 🡆 Linien ergeben d. Pfad, Werte im Pfad multiplizieren, zwischen d. Pfaden addieren (Bsp. rote, blaue grüne Kugel oder Schwarzfahrer)

Mann sucht Traumfrau mit best. Eigenschaften. prinzipiell auch wie Baum-Diagramm, nur mit einem Pfad: alle Wahrscheinlichkeiten werden multipliziert

**Siebformel:** *Wie groß ist die Wahrscheinlichkeit bei 2 (aufeinanderfolgenden) Würfen mindestens eine 5 zu erzielen?*

**(𝐀 + 𝐁) – (𝐀 ∩ B)**

**Gegenwahrscheinlichk***.: Wahrscheinlichkeit bei 6 (aufeinanderfolgenden) Würfen wenigstens einmal eine 6 zu erzielen?*P( A̅ ) =1 – P( A ) bzw. P( A ) =1 – P( A̅ ) 5 = „keine 5“, 6 = alle Zahlen

*Wahrscheinlichkeit, aus einem Schachspiel eine beliebige Figur genommen wird, die kein Pferd ist.*

Figuren im Spiel n = 32, n(Pferde) = 4 🡆 P(Pferd) = 4:32, P(keinPferd) = 1 – 4:32

*2 x eine Kugel mit Zurücklegen aus einer Urne. Wahrscheinlichk. für Ziehen mind. einer blauen Kugel* ***95/144****. W. für rote Kugel:*

Wurzel, weil P(rote Kugel im 1.Zug) \* P(rote Kugel im 1.Zug) = p²

**gleichz. Ziehen v. 3 Pack aus 20 Pack**; **5 Pack sind 2. Wahl; genau 1 beschäd. Pack**, o. Reihenf, o. Zurückl: KoW = n =20, k = 3 🡆 für |E| **n = 5** (2. Wahl), **k = 1** (besch. Pack); 2 Mengen

**höchstens 1:**  **mind. 1**

**Diagramme**  
WICHTIG: sprechende Achsenbeschriftungen, sinnvolle Skaleneinteilung, ggf. Legende

**Säulendiagramm**   
- höhenproportionale Häufigkeitsverteilung (senkrechte nicht aneinandergrenzende Säulen); Säulen können beliebig breit sein; y-Achse zur abs. Häufigkeit hi, x-Achse zu x-Werten, für wenige Ausprägungen

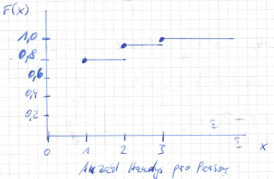
**Stabdiagramm** (Liniendiagramm) 🡆 wie Säulendiagramm nur mit schmalen Säulen

**Balkendiagramm** 🡆 am häufigsten verwendet, wie Säulendiagramm nur mit horizontalen Balken ( y-Achse zu x-Werten; x-Achse zur abs. Häufigkeit hi)

**Kreisdiagramm** 🡆 nur für eine Datenreihe, keine negativen Werte, keine 0-Werte, Kategorien repräsentieren Teile des gesamten Kreisdiagramms, max. 7 Teilwerte

**Histogramm**   
- **flächenproportionale Darstellung der absoluten oder relativen** (beide sind möglich) **Häufigkeiten** von ausschließlich klassierten Daten   
- **x-Achse**: Werte müssen auf Skala geordnet sein und gleiche Abstände haben,   
 x-Achse enthält **aneinandergrenzende Rechtecke zur Klassenbereite**, **keine Abstände zwischen den Flächen**

- **y-Achse**: **Rechteckhöhe** **Höhe = Häufigkeitskoeffizient** berechnet aus (**ri = hi / bi** (Klassenbreite)) oder (**fi / bi** ),   
 Skaleneinteilung entsprechend den berechneten Werten, **in den Rechtecken den Wert zur relativen Häufigkeit fi** eintragen

**empirische Verteilungsfunktion** (Treppenfunktion) F(x) ist (**relative) Summenhäufigkeitskurve**

- **x-Achse enthält x-Werte oder Klassengrenzen**,

- bei klassierten Daten werden die Punkte / Striche im Diagramm bei K-Obergrenze gezeichnet

- y-Achse enthält Skala zur relativen Summen-Häufigkeit Fi:;

- **Punkt** oder Strich (oder Punkt mit Strich) wird **zur Summenhäufigkeit auf y-Achse** **und**

**zum x-Wert bzw. zur K-Obergrenze auf x-Achse** gezeichnet

**Streudiagramm**

graphische Darstellung zur Abhängigkeit (Regression) und zum Zusammenhang (Korrelation) von **beobachteten Wertepaaren zweier Merkmale X, Y**  
Wertepaare werden in kartesisches Koordinatensystem eingetragen 🡆 ergibt Punktwolke  
Stärke und Richtung des Zusammenhangs anhand der Lage und Form der Punktwolke;   
erste Hinweise über mögliche Abhängigkeit zwischen Merkmalen

**Boxplot**

**W-Graph** (Baumdiagramm)