Lageparameter: x_D (Modus, Modalwert), x_Z(Median, Zentralwert, $\bar{x}$ (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): $X_D > X_Z > \bar{x}$, **rechtsschief** (linkssteil): $X_D < X_Z < \bar{x}$, metrische Verteilung $x_Z = \bar{x} = x_D$

2-dimensionale Häufigkeitsverteilung ➡ Kreuztabelle
M = Merkmal (A oder B), G = Geschlecht

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%

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)

Merkmals-ausprä-gung Anzahl der Handys	absolute Häufigk. Anz. Handy -Nutzer		relative Häufigk. h_i / n	abs. Summen-häufigk. $h_i + h_{i+1}$	rel. Summen-häufigk. $f_i + f_{i+1}$	arithm. Mittel $\sum(x_i * h_i)$ / n 25 / 20	Zwischen-rechnung für Varianz
x_i	h_i	$x_i * h_i$	f_i	H_i	F_i	$\bar{x}$	$(x_i - \bar{x})^2 * h_i$
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**

für klassierte Daten
$r_i = \frac{h_i}{b_i}$

$h_i = r_i * b_i$
$n = \sum h_i = h_1 + h_2 + h_3$
$20 = 16 + h_2 + h_3$ | | -16
$4 = h_2 + h_3$ | - h_3
$h_2 = 4 - h_3$
$\bar{x} = \frac{1}{n} \sum(x_i * h_i)$
$1{,}25 = \frac{1}{20} * 16 + 2h_2 + 3h_3$ | $\frac{1}{20}$
$25 = 16 + 2 * (4 - h_3) + h_3$ | - 16
$9 = 8 - 2h_3 + h_3$ | - 8
$1 = -2h_3 + h_3$ ➡ $h_3 = 1$
$h_2 = 4 - h_3 = 4 - 1 = 3$

Mittelwert $\bar{x}$ $\quad \bar{x} = \frac{1}{n} \sum x_i * h_i$

Mittelwert $\bar{x}$ bei klassierten Daten wie bei unklassierten Daten nur mit Klassenmitte m_i,

$\bar{x} = \frac{1}{n} \sum_i^n m_i * h_i$ $\qquad$ m_i **(Klassenmitte: $= \frac{1}{2} . (x_{k-1} + x_k) * h_i$**

Klassenbreite $b_i = x_k - x_{k-1}$ Bsp. Klasse „20 b. u. 35 ": Klassenbreite b_i = 35 – 20 = 15

bei klassierten Daten werden $\bar{x}$, s², s, v wie bei unklassierten Daten im Casio berechnet . Statt x_i wird m_i (Klassenmitte) in Tabelle eingetragen

Varianz s² **auf Grund der Quadrierung immer ≥ 0**
arithmetische Mittel der Abweichungsquadrate
wichtiger Streuungsparameter, f metrische Merkmale, Ausgangswert f. Standardabweichung & Variationskoeffi

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

$s^2 = 1/n * ((x_1 - \bar{x})^2 * h_1 + (x_2 - \bar{x})^2 * h_2 + (x_3 - \bar{x})^2 * h_3)$
$s^2 = 1/20 * ((1 - 1{,}25)^2 * 16 + (2 - 1{,}25)^2 * 3 + (3 - 1{,}25)^2 * 1)$

Varianz s² bei klassierten Daten

$s^2 = \frac{1}{n} \sum_{i=1}^{n} (m_i - \bar{x})^2 * h_i$ $\qquad$ wie bei unklassierten Daten nur mit Klassenmitte m_i statt mit x_i

Standardabweichung (durchschn. Abweichung, Standardfehler) s ➡ Streuung in einer Stichprobe

$s = \sqrt{s^2} = \sqrt{0{,}2875} = 0{,}536$

Varianzkoeffizient v
v = s / $\bar{x}$ **Standardabweichung s / Mittelwert $\bar{x}$**
Relatives Streuungsmaß (Streuungsparameter), dimensionslose Größe, **prozentuales Verhältnis der Standardabweichung zum arithmetischen Mittel, zum Vergleich der Streuung zwischen verschiedenen Erhebungen**

Modus $\bar{x}_D$ ➡ Beobachtungswert mit der größten Häufigkeit ➡ mehrere Merkmalsausprägungen mit der gleichen maximalen Häufigkeit = mehrere Modalwerte; Beobachtungsw. mit Häufigkeit == 1 ➡ kein Modus!) ➡ Modus x_D aus o. g. Tabelle ist 1
bei klassierten Daten ist Modalwert die Mitte der Klasse mit den größten Häufigkeiten (Modalklasse), Modalwert = m_i

Spannweite w
$w = x_{max} - x_{min}$ **auch bei klassierten Daten!**

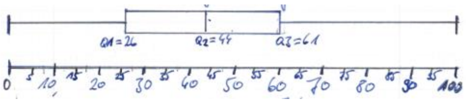
Quartile ➡ Q1, Q2, Q3 ➡ Q_2-Quartil = **Median / Zentralwert $\bar{x}_Z$** **Beobachtungs-Werte müssen geordnet sein!**

Median x_Z: mind. **50% der Beobachtungswerte (n) liegen links und rechts des Medians $X_Z = X_{n+1*1/2}$**
wenn n ungerade, z. B. n = 21 ➡ $X_{21+1 * ½} = x_{11}$ wenn n gerade, z. B. = 20 ➡ $(X_{20 * ½} + X_{20 * ½ + 1}) = (x_{10} + x_{11}) * 1/2$

Interquartilsabstand IQR = Q3 – Q1
Streuungsmaß, Differenz zwischen oberem und unterem Quartil enthält 50% der Verteilung.

Boxplot: enthält **x_{min}, Q1, Median, Q3, x_{max}**
informiert über Streuungsmaße Spannweite ($x_{max} - x_{min}$), IQR
Schiefe und Ausreiße
x_{min} und x_{max}, auch bei klassierten Daten

Achsenbeschriftung und Legende („mit Migrationshintergr.") nicht vergessen, beides unterhalb das Boxplots"
Achse enthält die Skala zu den x-Werten.



Median bei klassierten Daten

$$\bar{x}_Z = x_{k-1} + (x_k - x_{k-1}) * \frac{(0{,}5 - F_{k-1})}{f_k}$$

x_{k-1} = K-Untergrenze, x_k = K-Obergrenze
bei Q1, Q3 wird genauso verfahren.
Die Einfallsklasse muss rel. Summenhäufigkeit $F_i \geq 25\%$ (bei Q1) bzw. $F_i \geq 75\%$ (bei Q3) haben.
Für den %-Wert im Zähler des Bruchs wird statt 0,5 ➡ 0,25 (für Q1) bzw. 0,75 (für Q3) genommen

k = Einfallsklasse ➡ Klasse mit rel. Summenhäufigkeit $F_i \geq 50\%$,
x_{k-1} = Untergrenze d. Einfallsklasse, x_k = Obergrenze d. Einfallsklasse,
F_{k-1} = relative Summenhäufigkeit der Klasse unterhalb (vor) der Einfallsklasse
f_k = relative Häufigkeit der Einfallsklasse

x_i	h_i	f_i	H_i	F_i
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 k_2 „30 b. u. 40", da $F_i \geq 50$ (in Klasse „10 b. u. 30" ist $F_i < 50$
Hinweis. %-Werte immer als Dezimalwerte schreiben

$$\bar{x}_Z = 30 + (40 - 30) * \frac{(0{,}5 - 0{,}3)}{0{,}4} = 30 + 10 * \frac{(0{,}5 - 0{,}3)}{0{,}4}$$

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 $r_{x,y}$: Zusammenhangsmaß (Assoziationsmaß) für Quantifizierung der Korrelation $-1 \leq r_{xy} \leq +1$
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.)
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 r_{xy}: ➡ Wertebereich: $-1 \leq r_{xy} \leq +1$
- 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

$$r_{x,y} = \frac{\left(\frac{1}{n}\sum x_i * y_i\right) - \bar{x} * \bar{y}}{\sqrt{\left(\frac{1}{n}\sum x_i^2\right) - \bar{x}^2} * \sqrt{\left(\frac{1}{n}\sum y_i^2\right) - \bar{y}^2}}$$

$$r_{x,y} = \frac{\left(\frac{1}{4} * 450\right) - 4 * 25}{\sqrt{\left(\frac{1}{4} * 74\right) - 16} * \sqrt{\left(\frac{1}{4} * 3000\right) - 625}}$$

für Kovarianz cov(x, y) $\left(\frac{1}{n} * \sum x * y\right) - (\bar{x} * \bar{y})$ **im Zähler**
$\left(\frac{1}{n}\sum x_i * y_i\right)$ bezieht sich auf die Summe des zuvor zu ($x_i * y_i$) berechneten Produkts (**nicht die Summen von x und y multiplizieren, sondern das in jeder Tabellen-Zeile berechnete Produkt aus ($x_i * y_i$) addieren (hier 450) !**
für Standardabweichung s im Nenner
gilt dasselbe. Quadrate für x_i, y_i summieren (= 74 und 3000) und $\bar{x}^2$ bzw. $\bar{y}^2$ subtrahieren.

Bsp:
Werte in Tabelle erfassen:
x_i, y_i & $x_i * y_i$ & x_i^2, y_i^2 & Summen zu allen Werten & Mittelwert zu x_i und y_i & Quadrate zu Mittelwerten

Filial-Nr i	x_i VK-Fläche Tsd. m²	y_i Umsatz (Mio €)	x_i^2	y_i^2	$x_i * y_i$
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

$\bar{x}$ = 4 (16 : 4) $\bar{y}$ = 25 (100 : 4)
$\bar{x}^2$ = 16 $\bar{y}^2$ = 625

$$r_{x,y} = \frac{\left(\frac{1}{n}\sum x_i * y_i\right) - \bar{x} * \bar{y}}{\sqrt{\left(\frac{1}{n}\sum x_i^2\right) - \bar{x}^2} * \sqrt{\left(\frac{1}{n}\sum y_i^2\right) - \bar{y}^2}}$$

$$r_{x,y} = \frac{\left(\frac{1}{4} * 450\right) - 4 * 25}{\sqrt{\left(\frac{1}{4} * 74\right) - 16} * \sqrt{\left(\frac{1}{4} * 3000\right) - 625}} = \frac{112{,}5 - 100}{\sqrt{18{,}5 - 16} * \sqrt{750 - 625}}$$

$$= \frac{12{,}5}{\sqrt{2{,}5} * \sqrt{125}} = \frac{12{,}5}{\sqrt{1{,}581} * \sqrt{11{,}180}} = \frac{12{,}5}{17{,}676} = 0{,}707$$

Interpretation d. Korrelationskoeffizienten r_{xy}: nur Orientierungswerte → Entscheidung 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

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 (unabh. Merkmal); **Zweck: Erstellen v. Vorhersagemodelle, Quantifizierung d. Stärke d. Zusammenhangs**

Bestimmtheitsmaß $R^2 = (r_{x,y})^2$ ➡ immer in % angeben! (Quadrat der Korrelationskoeffizienten r in %) **immer ≥ 0**
0% ≤ R² ≤ 100% (bzw. 0 ≤ R² ≤ 1) aus o. g. Bsp: $r_{x,y}$ = 0,707 ➡ $R^2 = (r_{x,y})^2 = 0{,}707^2 = 0{,}50 = 50\%$
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, dann R² nahe 0

Regressionsfunktion (Dependenzanalyse)

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

Regressionsfunktion ŷ = a + b * x (nach **Methode der kleinsten Quadrate**) ➡ **x = (ŷ – a) / b**

Regressionskoeffizienten a, b ➡ linearer **Zusammenhang durch Gerade in Punkwolke**

Summe der quadratischen Abweichungen von a und b (Kurvenparameter) soll zu beobachteten Punkten minimal sein

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

$$a = \frac{\sum x_i^2 * \sum y_i - \sum x_i * \sum (x_i * y_i)}{n * \sum x_i^2 - (\sum x_i)^2} \qquad b = \frac{n * \sum (x_i * y_i) - \sum x_i * \sum y_i}{n * \sum x_i^2 - (\sum x_i)^2}$$

Formel im Nenner ist in beiden Gleichungen identisch

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

Filialen-Nr. i		x_i VK-fläche (1000 qm)	y_i Umsatz (Mio €)	x_i^2	y_i^2	$x_i * y_i$	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

Interpretation der R-Koeffizienten a und b:
a ist konstante Größe und ist unabhängig von Einflussgröße,
b ist variabler Faktor im veränderl. Term der R-Rechnung, b ist abhängig von Einflussgröße (unabhängiger Variable) x
ŷ erhöht/vermindert sich um {Wert d. R-Koeffiz. a}
ŷ = a + b * x **ŷ = 5 + 5 * x**

$$a = \frac{74 * 100 - 16 * 450}{4 * 74 - 16^2} = \frac{7400 - 7200}{296 - 256} = \frac{200}{40} = 5 \qquad b = \frac{4 * 450 - 16 * 100}{4 * 74 - 16^2} = \frac{1800 - 1600}{296 - 256} = \frac{200}{40} = 5$$

Varianz d. Residuen: $s_{\hat{y}}^2 = (\frac{1}{n} * \sum \hat{y}_i^2) - \bar{\hat{y}}^2$ in Hilfstabelle 1 Spalte mit werten aus **ŷ = a + b * x** ➡ Quadrate jedes **ŷ** addieren und durch n dividieren; davon Quadrat zum Mittelwert von $\bar{\hat{y}}^2$ 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

Wahrscheinlichkeitsrechnung

(A U B) U C = A U (B U C) <=> (A ∩ B) ∩ C = A ∩ (B ∩ C) A U B = B U A <=> A ∩ B = B ∩ A

A U (B ∩ C) = (A U B) ∩ (A U C) <=> A ∩ (B U C) = (A ∩ B) U (A ∩ C) A \ B = A ∩ B^C

Mengenoperationen

A U B = Vereinigungsmenge ➡ Objekte in **A ODER B** ➡ **(x ∈ A) v (x ∈ B)** ➡ **Addition**

A ∩ B = Schnittmenge ➡ Objekte in **A UND B** ➡ **(x ∈ A) v (x ∈ B)** ➡ **Multiplikation**

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

B^C oder B̄ 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)

Multiplikationsregel der Kombinatorik: mehrfache Auswahl ➡ mögliche Fälle $m_1 * m_2 * m_3$

z. B. 100 Tonfiguren, 20% fehlerhaft, P für 4 fehlerfreie Figuren

$m_1 = \frac{80}{100}$ (80% fehlerfrei, 20% fehlerhaft, 100 gesamt), $m_2 = \frac{79}{99}$ usw. $\frac{80}{100} * \frac{79}{99} * \frac{78}{98} * \frac{77}{97} = 0{,}40$

P. zur Los-Auswahl v. 5 Angriffsspielern bei 10 Spielern: $P(A_i) = \frac{k_i}{n_i}$ k_i *ist* E, n_i *ist* Ω $P = (\frac{5}{10} * \frac{4}{9} * \frac{3}{8} * \frac{2}{7} * \frac{1}{6})$

Permutat. m Wiederholung; Anordnung der k-Elemente ➡ Schlüsselwörter „Anzahl“ „Wie viele“

Anzahl d. Möglichkeiten zur Anordnung von Buchstaben oder Anzahl der Kombis f. Perlen auf Kette

$\frac{n!}{n_1! * n_2! * \ldots * n_k!}$ **ODER im Casio 6 + SHIFT + Taste "X" + 3 ==> 6 P3**

ELEVEN = **6 Buchstaben, 4 B-Typen** ➡ **3 x „E“, 1 x „L“, 1 x „V“, 1 x „N“** = $\binom{6}{3}$ **ODER** $\frac{6!}{3!}$ (Fakultät 1 fällt raus)

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

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

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

„E“ und „N“ am Beginn/Ende v. (E)**LEVE**(N) nicht berücksichtigen ➡ „**LEVE**“ ➡ **n= 4 und k = 2 (2 x „E“)** = $\binom{4}{2}$ ODER $\frac{4!}{2!}$;

bedingte Wahrscheinlichk. $P(A|B)$ ➡ Wahrscheinlichk für **Ereignis A nach Ereignis B** (**B ist d. 1. Bedingung im Graph**)

„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|B̄) * P(B̄) ➡ **P(p) = P(p|M) * P(M) + P(p|F) * P(F)** (im Pfad multiplizieren)

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

$$P(B_i|A) = \frac{P(A|B_i) * P(B_i)}{\sum_{i=1}^{n} P(A|B_i) * P(B_i)}, \qquad P(F|s) = \frac{P(s|F) * P(F)}{P(s|F) * P(F) + P(s|M) * P(M)}$$

bedingte W.: 4 rote, 6 blaue, zweimal Ziehen ➡ ist 1. Ball blau wird zurückgelegt, ist 2 Ball rot, wird nicht zurückgelegt

1. Kugel blau? ➡ P(blaue im 1.Zug) = 6/10; P(rote im 1.Zug) = 4/10; P(blaue im 2.Zug) = 6/10; P(rote im 2.Zug) = 4/10

1. Kugel rot? ➡ P(blaue im 1.Zug) = 6/10; P(rote im 1.Zug) = 4/10; P(blaue im 2.Zug) = 6/9; P(rote im 2.Zug) = 3/9

> blaue im 1. Zug und rote Kugel im 2. Zug: P(A)= 6/10*4/10 = 0,24

> 2 blaue: 6/10*6/10 + 6/10 * 6/9 = 0,76 > 2 rote: 4/10*4/10 + 4/10 * 3/9 = 0,29

mit Reihenfolge, mit Zurücklegen ➡ V_{mw} ➡ $|\Omega| = n^k$

4 x würfeln, P für verschiedene Zahlen? ➡ *n = 6, k = 4* $|\Omega| = 6^4$ **= 1296** $|E|$ = **Vow** = $\frac{6!}{(6-4)!}$ **= 360**

Wörter der Länge 3 aus 5 Buchstaben. P für Wörter mit 2 verschied. B? $|\Omega| = 3^5$ **= 125** $|E|$ = $\frac{5!}{(5-3)!}$ **= 60**

3 rote, 4 bl., 5 grü. Bälle ➡ **3 x Ziehen mit Zurückl, gefragt: n f. verschieden. Farben n=12, k=3** $\Omega=12^3$ **E=3 * 4 * 5 = 60**
3 rote, 4 bl., 5 gr Bälle ➡ **3 x Ziehen mit Zurückl, gefragt: n für gleiche Farben** $\Omega=12^3$ $E=3^3+4^3+5^3=216$
bei verschiedenen Farben (= beliebige Kombi d. Möglichkeiten) ➡ **Multiplikation,**
bei gleichen Farben (keine Kombi, da sortiert (gefiltert) nach Farben) ➡ **Addition**

---

mit Reihenfolge, ohne Zurücklegen ➡ V_{ow} ➡ $|\Omega| = \binom{n}{k} * k!$ oder $\frac{n!}{(n-k)!}$ Bruch kürzen, falls n sehr groß

Sonderfall wenn n = k, dann |Ω| = n! Anzahl Mögl. 5 Leute auf 5 freie (unterscheidbare) Plätze zu verteilen? ➡ **Fakultät 5!**

4 x würfeln, W für verschiedene Zahlen? ➡ n = 6, k = 4 **|Ω| =** V_{mw} 6^4 **= 1296** **|E| = Vow** = $\frac{6!}{(6-4)!}$ **= 360**

100 Sportler, **1 Wettbewerb**, Anzahl d. Ausgänge f. 1. Platz, 2. Platz … **Ω=Vow** = $\frac{100!}{(100-3)!}$

mit Reihenfolge da Reihenf. d. Plätze gefragt. **ohne Wiederholung** da nur 1 Wettbewerb

---

ohne Reihenfolge, mit Zurücklegen ➡ K_{mw} ➡ $|\Omega| = \binom{n+k-1}{k}$ *oder* $\frac{(n+k-1)!}{k!*(n-1)!}$

10 Sportler in **3 Wettbewerben**, jeder Wettbewerb genau eine Sieger. Wie viele Arten für Verteilung d. Preise?

n = 10 (Sportler), k = 3 $|\Omega| = \text{Vow} = \binom{10+3-1}{3} = \binom{12}{3} = 220$ ohne Reihenf., weil Rangfolge d. Plätze nicht gefragt

---

ohne Reihenfolge, ohne Zurücklegen ➡ K_{ow} ➡ $\Omega = \binom{n}{k}$ *oder* $\frac{n!}{k!*(n-k)!}$ (**Binomialkoeffizient**); k Elemente aus Menge n

gleichz. Ziehen v. 3 Pack aus 20 Pack; 5 Pack sind 2. Wahl ➡ ohne Reihenfolge, ohne Zurücklegen ➡ $\binom{n}{k}$

genau 1 beschäd. Pack, n =20, k = 3 ➡ $\Omega = \binom{20}{3} = 1140$ ➡ **für |E|:** $n_2 = 5$ (2. Wahl), **k = 1** (besch. Pack);

$E = \binom{5}{1} * \binom{15}{2} = 525$ ➡ $P(A) = \frac{525}{1140} = 0{,}46$

höchstens 1 Pack: $E = \binom{5}{1} * \binom{15}{2} + \binom{5}{0} * \binom{15}{3} = 980$ ➡ $P(B) = \frac{980}{1140} = 0{,}86$

mind. 1 Pack: $E = \binom{5}{0} * \binom{15}{3} = 455$ ➡ **Gegenw.: 1 – keine** ➡ $P(C) = 1 - \frac{456}{1140} = 0{,}60$

nacheinander Ziehen mit Zurücklegen v. 3 Pack aus 20 Pack; 5 Pack sind 2. Wahl ➡ **mit Reihenfolge, mit Zurücklegen**
selbiger Algo, wie oben, jedoch mit n^k **UND zusätzlich k = 3 (Anzahl gezogener Pack) als Faktor**

$\Omega = 20^3 = 8000$ **genau 1 Packung:** $E = 3 * 5^1 * 15^2 = 3375$ $P(A) = \frac{3375}{8000} = 0{,}42$

höchstens. 1 Packung ➡ $E = 3 * 5^1 * 15^2 + 5^0 * 15^3 = 3375 + 3375 = 6750$ $P(B) = \frac{6750}{8000} = 0{,}84$

mind. 1 Packung ➡ Gegenwahrscheinlichk: 1 – keine ➡ $E = 5^0 * 15^3 = 3375$ ➡ $P(C) = 1 - \frac{3375}{8000} = 0{,}58$

3 Parteien A, B. C mit **jeweils 7 Mitglied.** soll **ein 6-köpfiger Ausschuss** gebildet werden. Wie viele Zusammensetzt, gesamt?

3 Parteien haben gesamt 7 * 3 = **21 Mitglieder** ➡ Varianten für 6 köpfigen Ausschuss **Ω** ($\binom{21}{6} = 54264$

Wie viele Ausschusszusammensetzungen, wenn mindestens ein A-Mitglied dabei sein soll?

Ω = 21 – „ohne A" ➡ Von Ω Anteil der |B| und |C|-Mitglieder (2 x 7 = 14) subtrahieren **|Ω| - |A| =** $\binom{21}{6} - \binom{14}{6}$ **= 51.261**

6 rote, 4 grüne, 5 gleichzeitig Ziehen ➡ K_{ow}, gefragt: 2 rote ➡ $|\Omega| = \binom{10}{5} = 252$ $|E| = \binom{6}{2} * \binom{4}{3} = 1\ 60$

3 rote, 4 bl, 5 gr ➡ 3 gleichz. Ziehen ➡ K_{ow}, gefragt: alle unterschiedl. Farbe: $|\Omega| = \binom{12}{3}$ $|E| = \binom{3}{3} * \binom{4}{3} * \binom{5}{3}$

3 rote, 4 bl, 5 gr ➡ 3 gleichz. Ziehen ➡ K_{ow}, gefragt: alle gleiche Farbe: $|\Omega| = \binom{12}{3}$ $|E| = \binom{3}{3} + \binom{4}{3} + \binom{5}{3}$

100 Figuren, 20% fehlerh, 80% fehlerfrei., 4 werden entn., W. für 4 fehlerfreie $\frac{80}{100} * \frac{79}{99} * \frac{78}{98} * \frac{77}{97}$ Multipl. da beliebige Kombi,

Wahrscheinlichk. für 3 fehlerfreie aus 4 entnommenen ➡ k_{oW} $|\Omega| = \binom{100}{4}$ $E = \binom{80}{3} * \binom{20}{1}$

100 Sportler, **1 Wettbewerb**, Anzahl d. Ausgänge f. die ersten 3 Plätze **Ω=Kow** = $\binom{100}{3}$

ohne Reihenfolge da keine Reihenf. d. Plätze gefragt (nur die ersten 3 Plätze). **ohne Wiederholung** da nur 1 Wettbewerb

Siebformel (Schlüsselwörter: „Anzahl" „Wie viele")
2 Mengen: A ∪ B = A + B − A ∩ B

Wie viele Ausschusszusammensetzungen für EINEN 6-köpfigen Ausschuss mit mind. ein A-Mitglied und ein B-Mitglied?

Ω − A + B − A ∩ B ➡ $\binom{21}{6} - \left(\binom{14}{6} + \binom{14}{6} - \binom{7}{6}\right) = \binom{21}{6} - 2 * \binom{14}{6} + \binom{7}{6} = 48.265$

*Wahrscheinlichkeit bei **2 (aufeinanderfolgenden) Würfen mindestens eine 5** zu erzielen?*

(A + B) – (A ∩ B) $P(\text{mindEine5In2Würfen}) = \frac{1}{6} + \frac{1}{6} - \left(\frac{1}{6} * \frac{1}{6}\right) = \frac{11}{36}$ gefragte Zahl ist egal, selbes Ergebnis m. 4 oder 3

3 Mengen: A ∪ B ∪ C = A + B +C – (A ∩ B) – (A ∩ C) – (B ∩ C) + (A ∩ B ∩ C)

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

Ω – A + B +C – (A ∩ B) – (A ∩ C) – (B ∩ C) + (A ∩ B ∩ C ➡ $\binom{21}{6} - 3 * \binom{14}{6} + 3 * \binom{7}{6} - 0 = 45.276$

Studenten mit versch. Kursen

Gegenwahrscheinlichk.: *Wahrscheinlichkeit bei 5 (aufeinanderfolgenden) Würfen wenigstens einmal eine 4 zu erzielen?*

$P(A) = 1 - P(\bar{A})$ ➡ $(\text{minEine4in5Würfen}) = 1 - \frac{5}{6} * \frac{5}{6} * \frac{5}{6} * \frac{5}{6} * \frac{5}{6}$ ➡ $\frac{5}{6}$ = Wahrscheinlichk. von 5:6 keine 4 zu würfeln,

Wahrscheinlichkeit kein Pferd im Schachspiel: Figuren im Spiel n = 32, n(Pferde) = 4 ➡ P(keinPferd) = 1 – 4:32

2 x 1 Kugel m. Zurückl. aus 1 Urne. W. für mind. eine blauen Kugel **95/144**. gefragt W. für rote Kugel $P = \sqrt{1 - \frac{95}{144}}$

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 h_i, x-Achse zu x-Werten, für wenige Ausprägungen

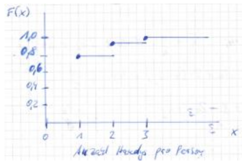
Stabdiagramm (Liniendiagramm) ➡ wie Säulendiagramm nur mit schmalen Säulen

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 ($r_i = h_i / b_i$ (Klassenbreite)) oder (f_i / b_i),
  Skaleneinteilung entsprechend den berechneten Werten, **in den Rechtecken den Wert zur relativen Häufigkeit f_i** 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 F_i:;
- **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)