Formelbuch WS

Grundbegriffe

- Statische Einheiten = $e \Rightarrow e_1, e_2, e_3, \dots e_i$
- Grundgesamtheit = $G = \{e_1, e_2, e_3, \dots, e_i, \dots e_n\}$
- Merkmal = $\alpha \quad (\alpha^{(1)}, \alpha^{(2)}, \alpha^{(3)}, \dots, \alpha^{(v)})$
- Statische Variable = $X^{(v)}$
- Ausprägung = $x_i^{(v)}$
- Skala = Ω Omega
- Skalenpunkt = ω Omega
- Klassen -> Faustregel $\approx \sqrt{n}$ Klassen
- Normklassenbreite = Die am häufigsten vertretene Klasse
- Dehnungsfaktore = $C_j = \dfrac{\text{Normalklassenbreite}}{\text{Breite der Klasse } K_j}$ $\quad K_j$ = Gefragte Klasse

Skalenniveau

Skalenniveau		Zulässige mathematische Operationen	Beispiele (Skala Ω und denkbare Ausprägungen ω)
Nominalskala		=/≠	Geschlecht {männlich, weiblich} Name {Anja, Alexander, Astrid, ...} Farbe {blau, rot, grün, ...}
Ordinalskala		=/≠; </> (auch im Sinne von „besser/schlechter" o.ä.)	Zufriedenheit {sehr unzufrieden, unzufrieden, indifferent, zufrieden, sehr zufrieden} Schulnoten {sehr gut, gut, ...} Dienstgrad {Jäger, Gefreiter, Obergefreiter, ...}
Metrisch (oder auch kardinal)	Intervallskala	=/≠; </>; +/−	Temperatur {..., −1, 0, 1, ...} °C Geburtsjahr {..., 2001, 2002, ...}
	Verhältnisskala	=/≠; </>; +/−; ·/÷	Monatseinkommen [0; ∞) EUR Alter [0; ∞) Jahre Gewicht [0, ∞) g Kinderzahl {0, 1, 2, ...}

Empirische Häufigkeitsfunktion

- Absolute empirische Häufigkeitsfunktion $\rightarrow f_h(x)$ (in reellen Zahlen)

  (Wie häufig ist etwas vorgekommen?)

$$f_h(x) = \begin{cases} n_k & \text{für } x = x_k \\ 0 & \text{sonst} \end{cases}$$

k = um welche Gruppe handelt es sich

nk = Gruppenhäufigkeit — z.B. erster Punkt 24-mal gemessen $\Rightarrow n_1 = 24$

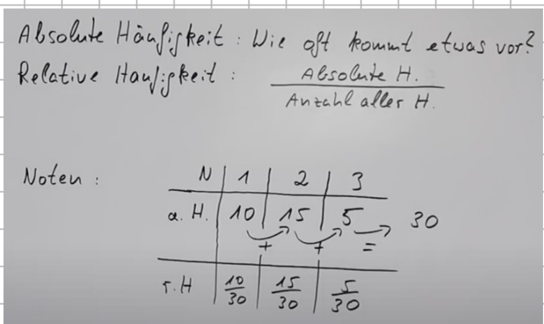
xk = Skalenwert — z.B. Gruppe k beinhaltet {schlecht, mittel, gut}, mit x_1 = schlecht, x_2 = mittel, x_3 = gut

x_1 = schlecht

- Relative empirische Häufigkeitsfunktion $\rightarrow f_h^*(x)$ (in Prozent)

$$\frac{\text{absolute Häufigkeit}}{\text{Anzahl aller Häufigkeiten}}$$

$$f_h^*(x) = \begin{cases} \frac{n_k}{n} = h_k^* & \text{für } x = x_k \\ 0 & \text{sonst} \end{cases}$$



- Empirische Häufigkeitsdichtefunktion

$$f_h(x) = \begin{cases} c_j \cdot n_j & \text{für } x \in [x_j ; x_{j+1}) \\ 0 & \text{sonst} \end{cases}$$

[= einschliesslich

) = ausschliesslich

[10, 30) = 10 - 29

Dichtefunktion $c_j = \dfrac{\text{Normalspaltenbreite}}{\text{Breite der Klasse } K_j}$

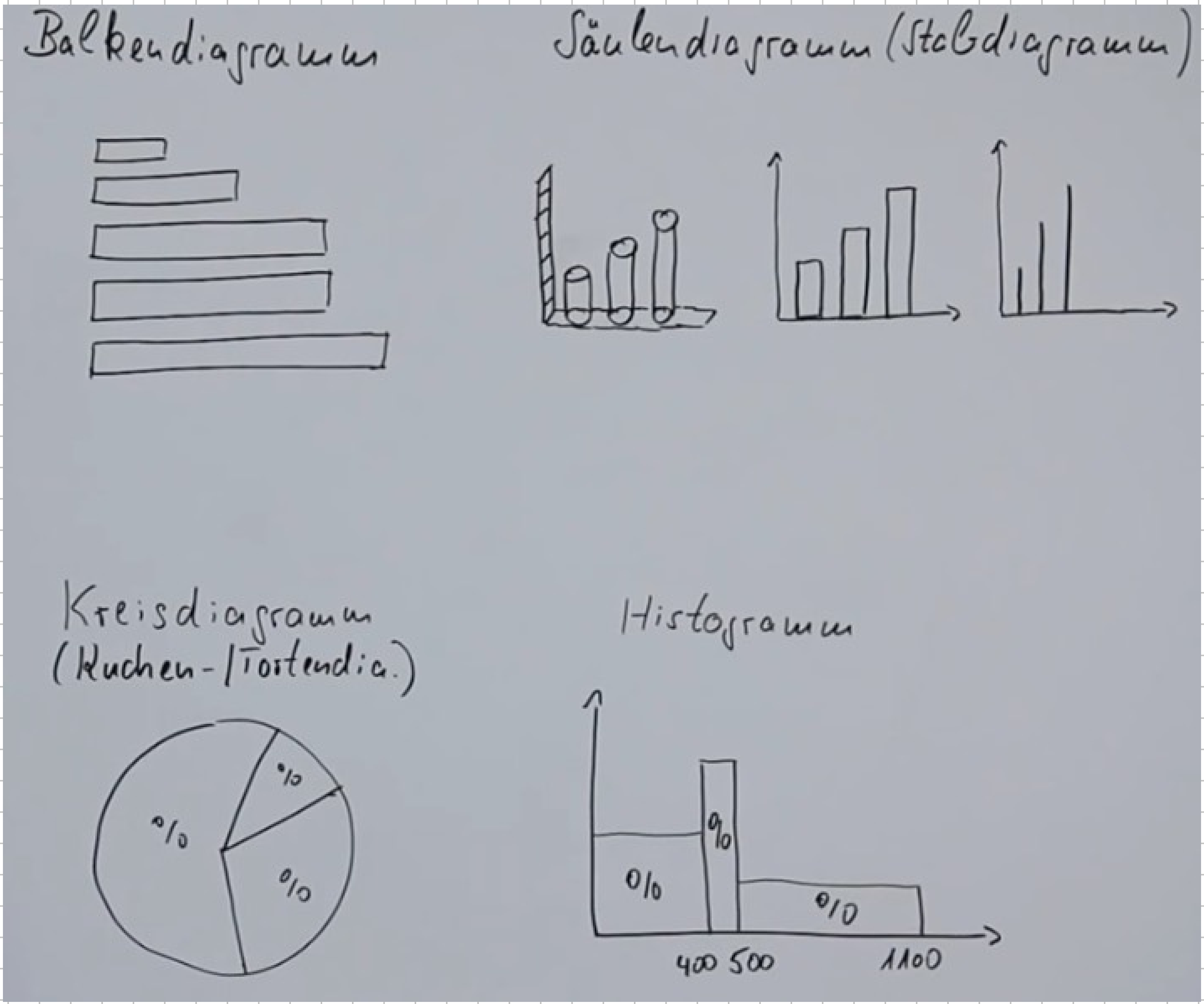
j = Um welche Klasse geht es

n_j = Häufigkeit der Klasse j — z.B. 24 Werte in K_1 = $n_1 = 24$

x_j = Untergrenze der Klasse j

x_{j+1} = Obergrenze der Klasse j

Übersicht Diagramme



Prozentberechnung für Winkel

Klasseneinteilung

Eignung statischer Darstellungsoptionen

Kontinuität	Skalenniveau	Wertetabelle	Wertediagramm	Häufigkeitstabelle	Stabdiagramm	Klassenhäufigkeitstabelle	Histogramm	Kreisdiagramm	Verteilungstabelle	Treppendiagramm
diskret	nominal	wenn n nicht zu groß ist	Piktogramme[1]	☺	😐	☹	☹	☺	☹	☹
	ordinal		wenn n nicht zu groß ist	☺	☺	☹	☹	😐	😐	😐
	metrisch (wenige verschiedene Messwerte)			☺	☺	😐	😐	😐	☺	☺
stetig quasi-stetig	metrisch			☹	☹	☺	☺	klassierte Daten	☺	☺

Beispiel: diskret

ω_k = Skalenpunkte

k → = Gruppen

Nummerierung k	Wert ω_k	absolute Häufigkeit n_k	relative Häufigkeit n_k^*	in %
1	nicht zufrieden	4	0,08	8
2	eher nicht zufrieden	5	0,10	10
3	neutral	15	0,29	29
4	eher zufrieden	17	0,33	33
5	zufrieden	10	0,20	20
		$n = 51$	1,00	100

n_k = absolute Gruppenhäufigkeit

n_k^* = relative Gruppenhäufigkeit

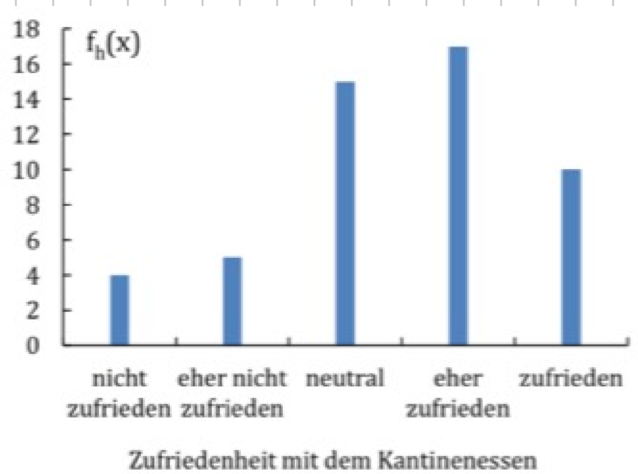
$$n = \sum_{k=1}^{5} n_k = n_1 + n_2 + n_3 + n_4 + n_5$$
$$= 4 + 5 + 15 + 17 + 10$$
$$= 51$$

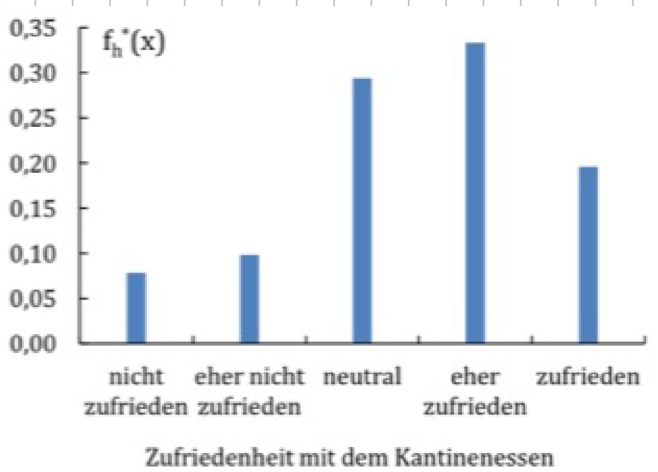
$$n^* = \sum_{i=1}^{5} \underbrace{\frac{n_k}{n}}_{=n_k^*} = \frac{n_1}{n} + \frac{n_2}{n} + \frac{n_3}{n} + \frac{n_4}{n} + \frac{n_5}{n}$$
$$= \frac{4}{51} + \frac{5}{51} + \frac{15}{51} + \frac{17}{51} + \frac{10}{51}$$

				%
$n_1^* = \frac{4}{51}$	$= 0{,}08$	$\cdot 100$		8
$n_2^* = \frac{5}{51}$	$= 0{,}10$	"		10
$n_3^* = \frac{15}{51}$	$= 0{,}29$	"		29
$n_4^* = \frac{17}{51}$	$= 0{,}33$	"		33
$n_5^* = \frac{10}{51}$	$= 0{,}21$	"		20
				100

$$f_h(x) = \begin{cases} 4 & \text{für "nicht zufrieden"} \\ 5 & \text{für "eher nicht zufrieden"} \\ 15 & \text{für "neutral"} \\ 17 & \text{für "eher zufrieden"} \\ 10 & \text{für "zufrieden"} \\ 0 & \text{sonst} \end{cases} \qquad f_h^*(x) = \begin{cases} 0{,}08 & \text{für "nicht zufrieden"} \\ 0{,}10 & \text{für "eher nicht zufrieden"} \\ 0{,}29 & \text{für "neutral"} \\ 0{,}33 & \text{für "eher zufrieden"} \\ 0{,}20 & \text{für "zufrieden"} \\ 0 & \text{sonst} \end{cases}$$

Stabdiagramm





Beispiel: stetig

- Klassierung:

n = Einheiten

m = Klassen

$m \approx \sqrt{n}$

z. B. 30 Unternehmen

$m \approx \sqrt{30} = 5$ oder $\boxed{6}$

Grösster Wert = 217,267 Mrd.

$217k / 6 \approx 36k$

Klasse	Intervall Mrd. EUR	absolute Klassen-häufigkeit n_j	relative Klassen-häufigkeit n_j^*
K_1	[0; 37)	15	0,5
K_2	[37; 74)	10	0,33
K_3	[74; 111)	2	0,1
K_4	[111; 148)	1	0,03
K_5	[148; 185)	1	0,03
K_6	[185; 222]	1	0,03
Σ		$n =$ 30	1

$n_k^* = \frac{n_k}{n} = \quad n_1 = \frac{15}{30} = 0,5$

$n_2 = \frac{10}{30} = 0,33$

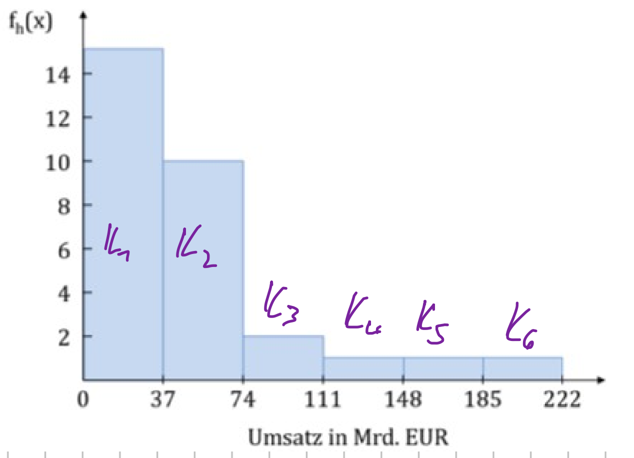
$n_3 = \frac{2}{30} = 0,1$

$n_4 = \frac{1}{30} = 0,03$

$n_5 = \frac{1}{30} = 0,03$

$n_6 = \frac{1}{30} = 0,03$

Histogramm



Unterschiedliche Klassenbreite, Fläche soll proportional zur Häufigkeit sein.

Normalklassenbreite = 25

Klasse	Intervall Mrd. EUR	absolute Klassen-häufigkeit n_j	Klassen-breite	Dehnungs-faktor c_j	$f_h(x) = c_j \cdot n_j$
K_1	[0; 25)	13	25	1	13
K_2	[25; 50)	9	25	1	9
K_3	[50; 75)	3	25	1	3
K_4	[75; 100)	2	25	1	2
K_5	[100; 150)	1	50	1/2	1/2
K_6	[150; 225]	2	75	1/3	2/3

$c_j = c_{1-4} = \frac{25}{25} = 1$

$c_5 = \frac{25}{50} = 1/2$

$c_6 = \frac{25}{75} = 1/3$

$f_h(x) = c_j \cdot n_j$

$x_1 = 1 \cdot 13$

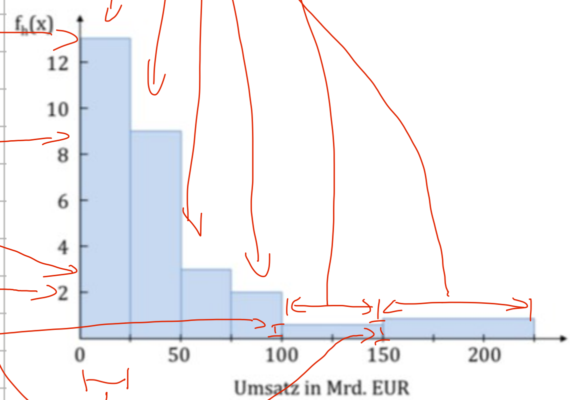
$x_2 = 1 \cdot 9$

$x_3 = 1 \cdot 3$

$x_4 = 1 \cdot 2$

$x_5 = 1/2 \cdot 1 = 1/2$

$x_6 = 1/3 \cdot 2 = 2/3$



Kreisdiagramm

- Absolute und daraus relative Häufigkeit bestimmen
- Relative Häufigkeit mit 360 multiplizieren
- Zu kleine Werte unter „ sonstiges“ zusammenfassen
- Der Grösste nach im Uhrzeigersinn ein Kreis zeichnen

k	Artikel ω_k	absolute Häufigkeit n_k	relative Häufigkeit n_k^*
1	Becher	12	0,24
2	Feuerzeug	1	0,02 ×
3	Flaschenöffner	2	0,04 ×
4	Jo-Jo	1	0,02 ×
5	LED-Lampe	18	0,36
6	Schlüsselanhänger	1	0,02 ×
7	Stift	2	0,04 ×
8	Tasche	1	0,02 ×
9	Textmarker	4	0,08
10	USB-Stick	8	0,16
×	Sonstige	$n = 50$	1,00

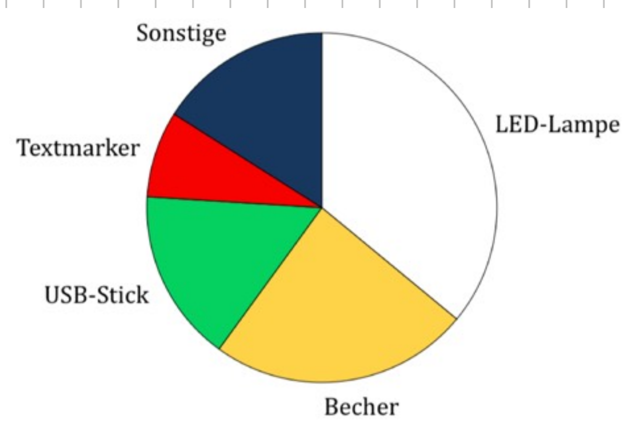
$$n = \sum_{k=1}^{\text{anzahl Einheiten}} n_k = 12+1+2+1+18+1+2+1+4+8 = 50$$

$$n_k^* = \frac{n_1}{n} = \frac{12}{50} = 0{,}24 \;,\; \frac{n_2}{n} = \frac{1}{50} = 0{,}02 \;, \ldots$$

$$\frac{n_5}{n} = \frac{18}{50} = 0{,}36 \;, \ldots$$

Anteil von k für den Kreis

k (neu)	Artikel ω_k	relative Häufigkeit n_k^*	Kreissektor $n_k^* \cdot 360°$
1	LED-Lampe	0,36	129,6°
2	Becher	0,24	86,4°
3	USB-Stick	0,16	57,6°
4	Textmarker	0,08	28,8°
5	Sonstige	0,16	57,6°
Σ		1,00	360°



Treppendiagramm

Absolute empirische Verteilungsfunktion

$$F_h(x) = \begin{cases} 0 & \text{für } x < x_1 \\ \sum\limits_{x_i \le x} f_h(x_i) & \text{für } x \in [x_1, x_n) \\ n & \text{für } x \ge x_n \end{cases}$$

→ x-Werte die kleiner als der kleinste Wert sind

→ x-Werte die aus dem Bereich der gemessenen Werte stammt, Summe aller Häufigkeiten aller Messwerte

→ x-Werte ≥ den grössten Messwert sind, die Häufigkeit n bekommen

Relative empirische Verteilungsfunktion

$$F_h^*(x) = \begin{cases} 0 & \text{für } x < x_1 \\ \frac{1}{n} \cdot \sum\limits_{x_i \le x} f_h(x_i) & \text{für } x \in [x_1, x_n) \\ n & \text{für } x \ge x_n \end{cases}$$

1. Messwerte aufsteigend ordnen

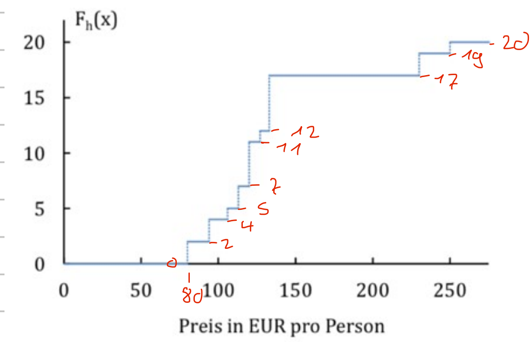
2. Häufigkeit der Messwerte bestimmen
   auch wenn es sich um stetige und quasi-stetige handelt

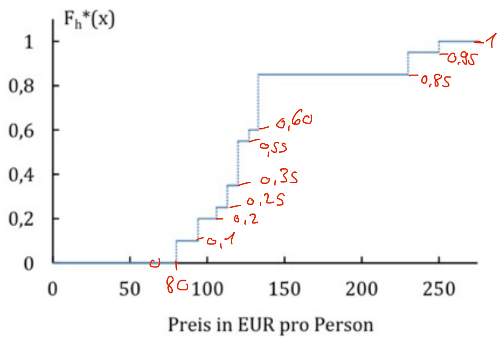
3. Aufsummieren aller Häufigkeiten bis zum jeweiligen Messwert

i	Preis x_i in EUR pro Person	absolute Häufigkeit n_i	absolute Summenhäufigkeit $\sum_{x_i \le x} f_h(x_i)$	relative Summenhäufigkeit $\frac{1}{n}\sum_{x_i \le x} f_h(x_i)$
1	80	2	2	0,10
2	94	2	4	0,20
3	106	1	5	0,25
4	113	2	7	0,35
5	120	4	11	0,55
6	127	1	12	0,60
7	133	5	17	0,85
8	230	2	19	0,95
9	250	1	20	1,00

2/20, 4/20, 5/20, …, 17/20, 19/20, 20/20

$$F_h(x) = \begin{cases} 0 & \text{für } x < 80 \\ 2 & \text{für } x \in [80; 94) \\ 4 & \text{für } x \in [94; 106) \\ 5 & \text{für } x \in [106; 113) \\ 7 & \text{für } x \in [113; 120) \\ 11 & \text{für } x \in [120; 127) \\ 12 & \text{für } x \in [127; 133) \\ 17 & \text{für } x \in [133; 230) \\ 19 & \text{für } x \in [230; 250) \\ 20 & \text{für } x \ge 250 \end{cases} \qquad F_h^*(x) = \begin{cases} 0 & \text{für } x < 80 \\ 0{,}10 & \text{für } x \in [80; 94) \\ 0{,}20 & \text{für } x \in [94; 106) \\ 0{,}25 & \text{für } x \in [106; 113) \\ 0{,}35 & \text{für } x \in [113; 120) \\ 0{,}55 & \text{für } x \in [120; 127) \\ 0{,}60 & \text{für } x \in [127; 133) \\ 0{,}85 & \text{für } x \in [133; 230) \\ 0{,}95 & \text{für } x \in [230; 250) \\ 1 & \text{für } x \ge 250 \end{cases}$$





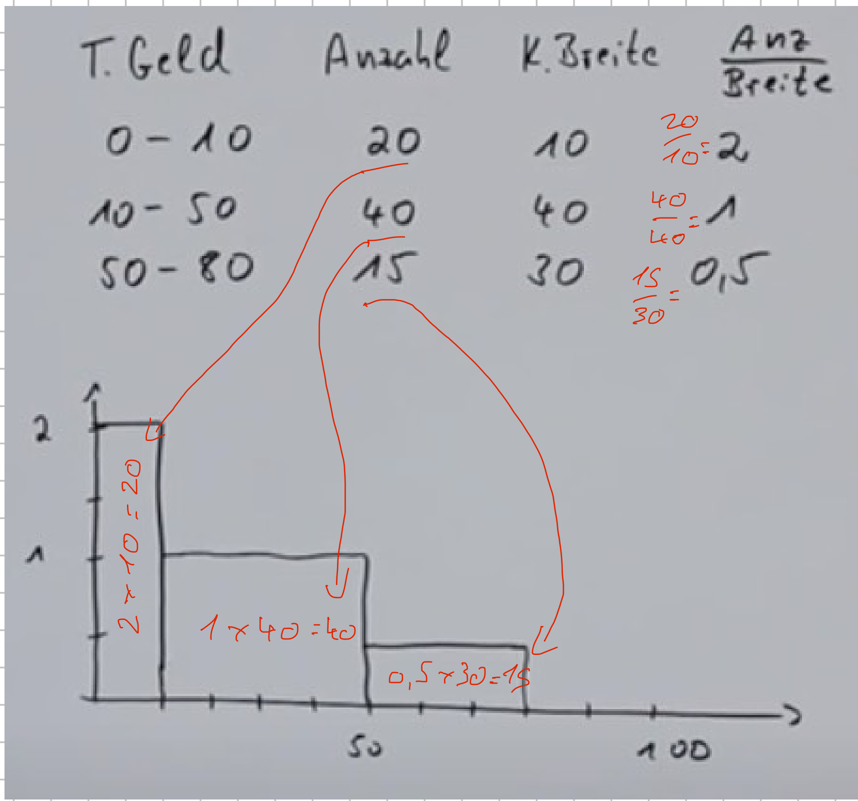
Histogramm:

Merkmal:

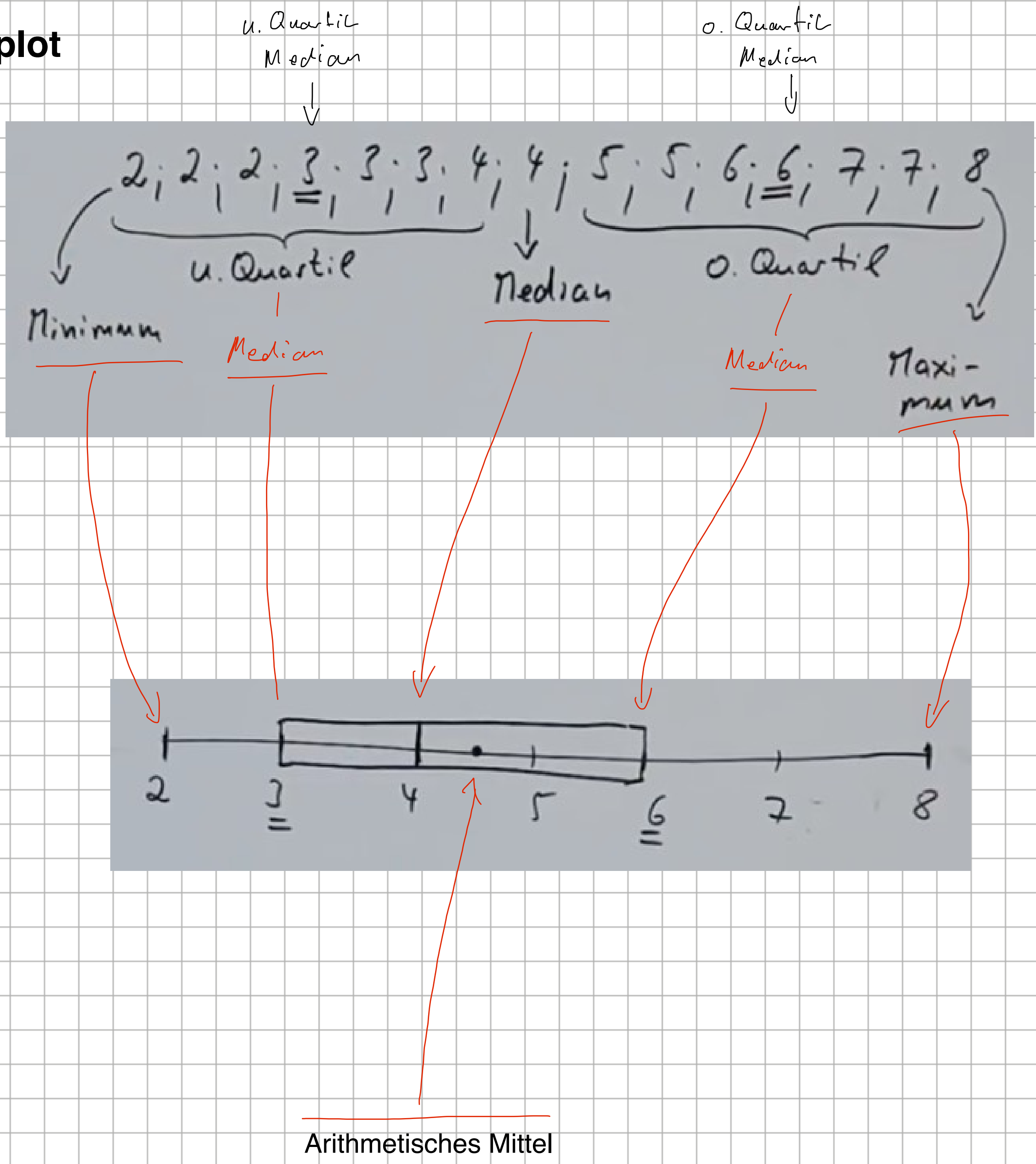
Merkmalsausprägung:

Merkmalsträger:

Grundgesamtheit:



Boxplot



$$\frac{3\cdot 2 + 3\cdot 3 + 2\cdot 4 + 2\cdot 5 + 2\cdot 6 + 2\cdot 7 + 8}{15} \approx 4{,}5$$

Quantil, Quartil, Dezil, Perzentile

1. Quartil: 0,25 - Quantil, 2. Quartil: 0,5 - Quantil (Median), 3. Quartil: 0,75 - Quantil

$\{x_1, x_2, \ldots, x_{16}, x_{17}\}$

$n_{0.25} = 0{,}25 \cdot 17 = 4{,}25 \notin \mathbb{Z} \Rightarrow X_{Q_{0.25}} = x_{[4,25]} \rightarrow x_{[5]}$

$n_{0.5} = 0{,}5 \cdot 17 = 8{,}5 \notin \mathbb{Z} \Rightarrow X_{Q_{0,5}} = x_{[8,5]} \rightarrow x_{[9]}$

$n_{0.75} = 0{,}75 \cdot 17 = 12{,}75 \notin \mathbb{Z} \Rightarrow X_{Q_{0,75}} = x_{[12,75]} \rightarrow x_{[13]}$

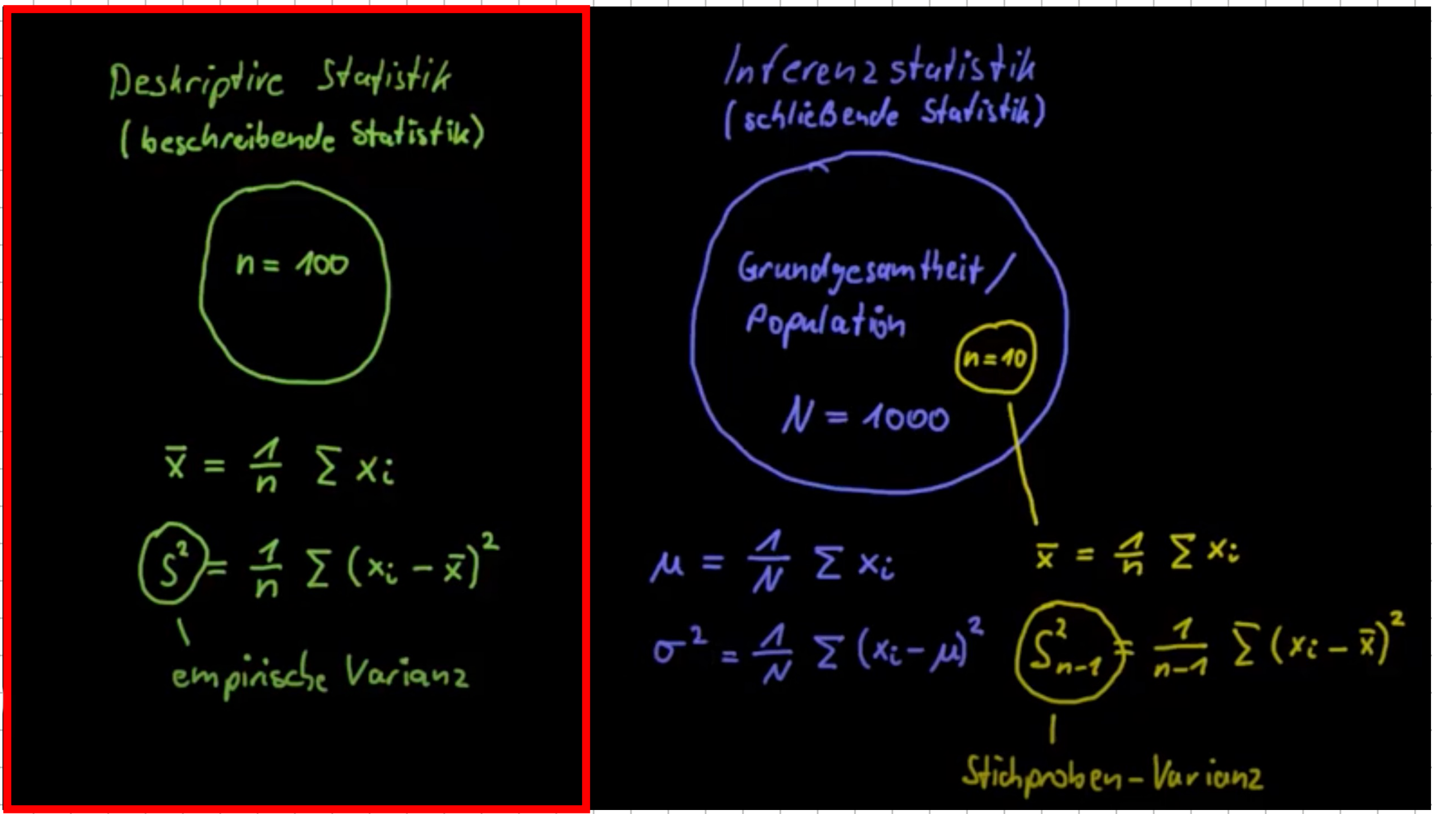
Interquantilabstand

3. Quartil - 1. Quartil

z. B.:

$x_{[13]} - x_{[5]}$

Deskriptive Statistik - Beschreibende Statistik



Empirisch -> Beachtung

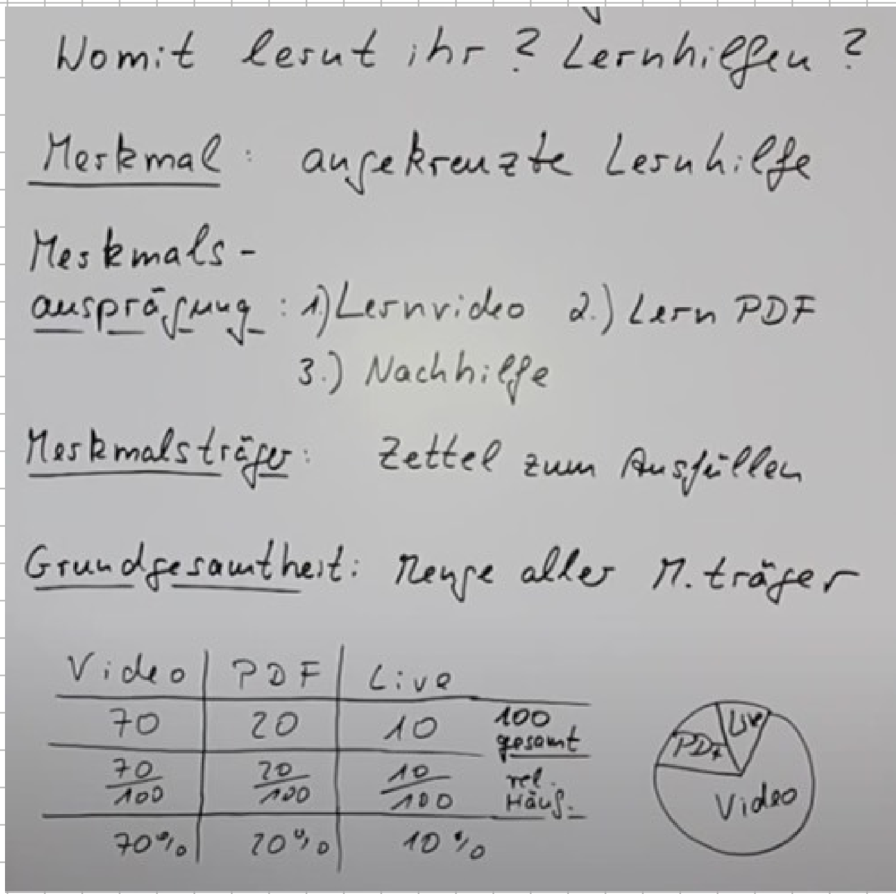
Statistische Erhebungen

Merkmal

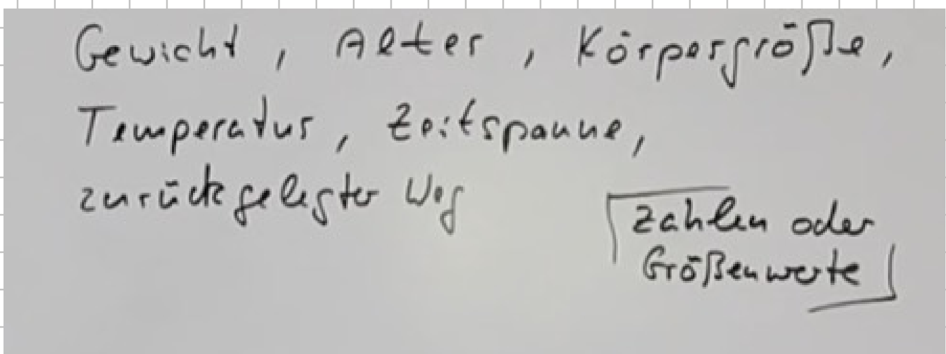
Merkmalsausprägung

Merkmalsträger

Grundgesamtheit

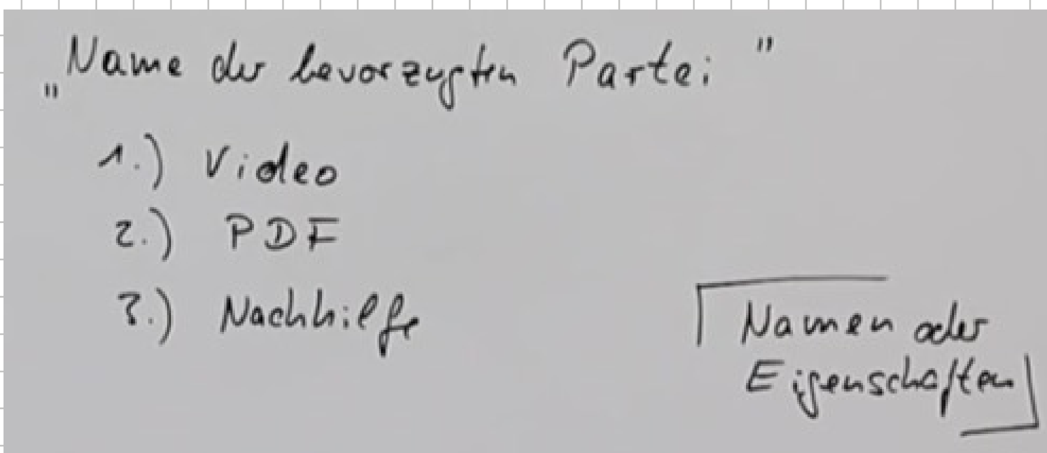


Ausprägungen bei: Quantitativen Merkmalen



Zahlen oder Grössenwerte

Ausprägungen bei: Qualitativen Merkmalen



Namen oder Eigenschaften

Masszahlen zur Beschreibung statischer Verteilungen

Mittelwerte:

Modalwert $= \bar{x}_{Mod}$ = Datensatz der am häufigsten vorkommt

$$f_h(\bar{x}_{Mod}) \geq f_h(x_i) \text{ für alle } i$$

$f_h(x_i)$ = absolute empirische Häufigkeit des i-ten Messwerts

$f_h(\bar{x}_{Mod})$ = Häufigkeit des Modalwertes, keiner darf häufiger vorkommen

Ermittlung

- bestimmung der Häufigkeiten
- den am häufigsten gemessenen Wert bestimmen

Median $= \bar{x}_Z$ oder $\bar{x}_{Med}$ = Genaue Mitte des Datensatzes

(Zentralwert)

$$\bar{x}_Z = \begin{cases} x_{[\frac{n+1}{2}]} & \text{falls } n \text{ ungerade ist} \\ \frac{1}{2}\left(x_{[\frac{n}{2}]} + x_{[\frac{n}{2}+1]}\right) & \text{falls } n \text{ gerade ist} \end{cases}$$

Eckige Klammer im Index:
* [ ] = geordneter Datensatz

Ermittlung

- Datensatz aufsteigend ordnen
- gerade oder ungerade anzahl Messwerte
- Median mit Hilfe Formel bestimmen

2, 2, 3, 4, 4, 5, 5, 6, 6

(Median ungerade)

$\frac{9 \text{ Werte} + 1}{2}$ = 5. Stelle $\Rightarrow \frac{n+1}{2}$

2, 2, 3, 4, 4, 5, 5, 6, 6, 7

$\frac{4+5}{2} = 4,5$

(Median gerade)

$\frac{10 \text{ Werte}}{2}$ = 5. Stelle, 6. Stelle

Note	1	2	3	4
Anz.	7	8	4	5

$= \frac{24 \text{ Werte}}{2}$ = 12., 13. } Wo liegen diese Werte?

7 + 8 → 15 = 12. = 2, 13. = 2

1, 1, 1, ..., 2 ....,

(Median aus Tabelle)

Mittelwerte

Mittelwerte

		Modalwert	Median	Arithmetisches Mittel	Geometrisches Mittel	Harmonisches Mittel
	Hauptsächliche Anwendung	bei nominal skalierten Merkmalen	bei ordinal skalierten Daten, aber auch bei metrischen Daten (z.B. wenn Randwerte bzw. Ausreißer keinen Einfluss haben sollen)	bei metrisch skalierten Datensätzen, wenn alle Werte gleichermaßen einfließen sollen	für Durchschnittsbildung über aufeinander aufbauenden Wachstumsraten	für Durchschnittsbildung über Verhältniszahlen, wenn die gegebene Bezugseinheit bzw. gegebene Gewichte sich auf den Zähler beziehen
	Symbol	$\bar{x}_{\text{Mod}}$	$\bar{x}_{\text{Z}}$	$\bar{x}$	$\bar{x}_{\text{g}}$	$\bar{x}_{\text{h}}$
Formel	**Allgemein**	$f_h(\bar{x}_{\text{Mod}}) \geq f_h(x_i)$ für alle i	$\begin{cases} x_{\left[\frac{n+1}{2}\right]} & \text{für ungerades } n \\ \frac{1}{2}\left(x_{\left[\frac{n}{2}\right]} + x_{\left[\frac{n}{2}+1\right]}\right) & \text{für geradesn } n \end{cases}$	$\frac{1}{n}\sum_{i=1}^{n} x_i$	$\sqrt[n]{\prod_{i=1}^{n} x_i}$	$\frac{n}{\sum_{i=1}^{n}\frac{1}{x_i}}$
	Gruppierte Daten	$\bar{x}_{\text{Mod}}$ ist der Wert der Gruppe mit der größten Häufigkeit	wie allgemeiner Fall	$\frac{1}{n}\sum_{k=1}^{r} n_k \cdot x_k$	$\sqrt[n]{\prod_{k=1}^{r} x_k^{n_k}}$	$\frac{n}{\sum_{k=1}^{r} n_k \cdot \frac{1}{x_k}}$
	Klassierte Daten	unüblich	wie allgemeiner Fall	$\approx \frac{1}{n}\sum_{j=1}^{m} n_j \cdot x_j'$	$\approx \sqrt[n]{\prod_{j=1}^{m} (x_j')^{n_j}}$	$\approx \frac{n}{\sum_{j=1}^{m} n_j \cdot \frac{1}{x_j'}}$
Zulässigkeit	**nominal**	☺	☹	☹	☹	☹
	ordinal	☺	☺	☹	☹	☹
	intervallskaliert	bei vielen Mehrfachnennungen	☺	☺	☹	☹
	verhältnisskaliert		☺	☺	☺	solange kein x_i (bzw. x_k, x_j') = 0

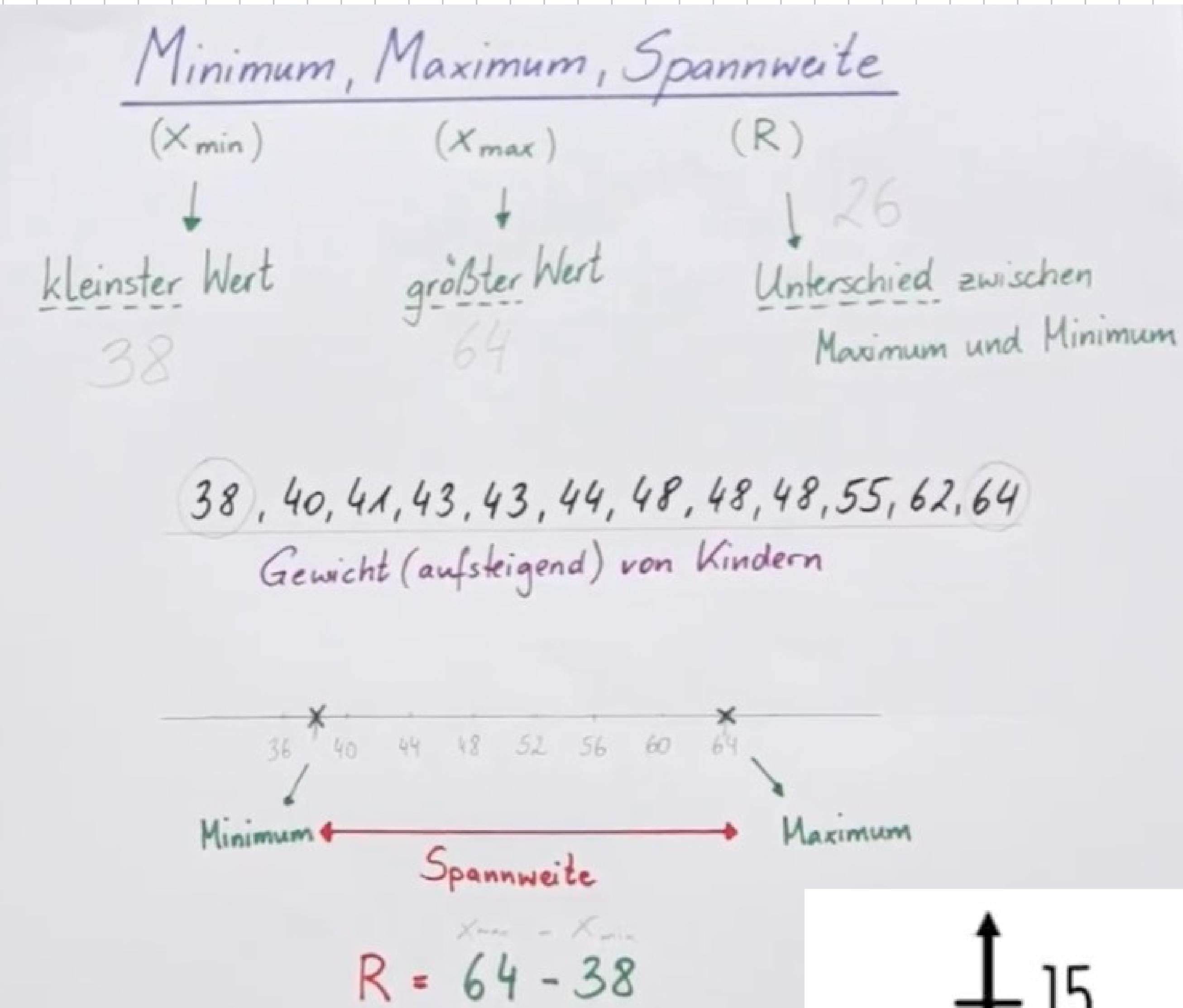
Streumasse (Streuparameter)

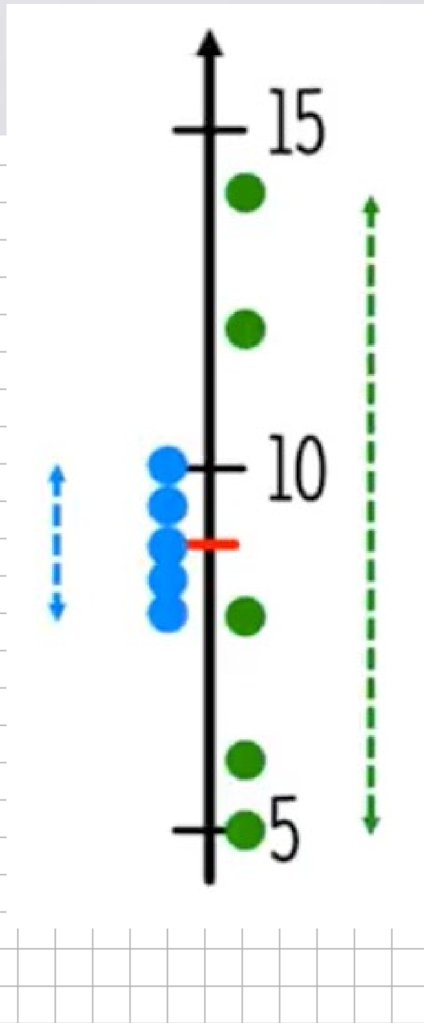
Streumaße

		Spannweite	Varianz	Standardabweichung	Variationskoeffizient	Quantil(-e)
	Symbol	R	s_x^2	s_x	v	$x_{Q[p]}$
Formel	**Allgemein**	$x_{[n]} - x_{[1]}$	$\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right) - \bar{x}^2$	$\sqrt{s_x^2}$	$\frac{\sqrt{s_x^2}}{\bar{x}} = \frac{s_x}{\bar{x}}$	Wert, der die aufsteigend geordnete Beobachtungsreihe im Verhältnis p zu $1-p$ teilt
	Gruppierte Daten	$x_r - x_1$	$\frac{1}{n}\sum_{k=1}^{r} n_k(x_k - \bar{x})^2 = \left(\frac{1}{n}\sum_{k=1}^{r} n_k x_k^2\right) - \bar{x}^2$	$\sqrt{s_x^2}$	$\frac{\sqrt{s_x^2}}{\bar{x}} = \frac{s_x}{\bar{x}}$	wie allgemeiner Fall
	Klassierte Daten	$\approx x_m' - x_1'$ (unüblich)	$\approx \frac{1}{n}\sum_{j=1}^{m} n_j (x_j' - \bar{x})^2$	$\approx \sqrt{s_x^2}$	$\approx \frac{\sqrt{s_x^2}}{\bar{x}} = \frac{s_x}{\bar{x}}$	wie allgemeiner Fall
Zulässigkeit	**nominal**	☹	☹	☹	☹	☹
	ordinal	☹	☹	☹	☹	☺
	intervallskaliert	☺	☺	☺	☺	☺
	verhältnisskaliert	☺	☺	☺	☺	☺

Tabelle 19: Zusammenfassung der Maßzahlen zu Lage- und Streuparametern (☺ - grundsätzlich geeignet; ☹ - ungeeignet)

Spannweite


Minimum, Maximum, Spannweite
(x_{min})
(x_{max})
(R)
kleinster Wert
38
größter Wert
64
26
Unterschied zwischen Maximum und Minimum
38, 40, 41, 43, 43, 44, 48, 48, 48, 55, 62, 64
Gewicht (aufsteigend) von Kindern
36 40 44 48 52 56 60 64
Minimum
Maximum
Spannweite
$x_{max} - x_{min}$
R = 64 - 38


15
10
5

Arithmetisches Mittel $= \bar{x}$ „Durchschnitt“

- Totalwert durch Anzahl Messwerte teilen
(Summe aller Messwerte)

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\left(x_1 + x_2 + x_3 + \ldots + x_n\right)$$

$$\bar{x} = \frac{1}{4} \cdot \left(110 + 210 + 190 + 50\right) = 140$$

x_i = die Messwerte

$\sum_{i=1}^{n} x_i$ = addiere alle Messwerte von x_1 bis x_n

n = anzahl aller Messwerte

$\frac{1}{n}$ = division durch anzahl Messwerte

Ges x

$$140 = \frac{110 + 210 + x + 50}{4} \quad | \cdot 4$$

$$4 \cdot 140 = 110 + 210 + x + 50 \quad | -110 \; | -210 \; | -50$$

$$4 \cdot 140 - 110 - 210 - 50 = x = \underline{\underline{190}}$$

Gewichtetes arithmetisches Mittel $= \bar{x}$

Gruppierte Daten, Messwerte mit Häufigkeit multiplizieren

$$\bar{x} = \frac{1}{n} \sum_{k=1}^{r} n_k \cdot x_k = \sum_{k=1}^{r} n_k^* \cdot x_k$$

Preis x_k in EUR/l	Bezogene Menge n_k in l
4,20	110
4,80	210
5,20	190
6,00	50
Gesamtmenge	**560**

$$\bar{x} = \frac{1}{560} \cdot \left(110 \cdot 4{,}20 + 210 \cdot 4{,}80 + 190 \cdot 5{,}20 + 50 \cdot 6{,}00\right) = \frac{2758}{560} = \underline{\underline{4{,}925}} \; \frac{EUR}{L}$$

Weitsprung: 7 m, 5 m, 7 m, 4 m, 8 m, 7 m, 4 m, 7 m, 4 m, 8 m (Urliste)

Sortieren: 4, 4, 4, 5, 7, 7, 7, 7, 8, 8 (Rangliste)

Tabelle:

Weite	4	5	7	8
Anzahl	3	1	4	2

Mittelwert: Arithmetisches Mittel

$$\frac{4+4+4+5+\ldots+8}{10} \quad \text{oder} \quad \frac{3 \cdot 4 + 1 \cdot 5 + 4 \cdot 7 + 2 8}{10}$$

Arithmetisches Mittel

gewichtetes Arithmetisches Mittel

„Durchschnittlicher Abstand zum Mittelwert“

Mittlere lineare Abweichung vom Mittelwert

(Mittlere absolute Abweichung)

$$d = \frac{1}{n} \cdot \left(|x_1 - \bar{x}| + |x_2 - \bar{x}| + \ldots + |x_n - \bar{x}| \right)$$

$$= \frac{1}{n} \cdot \sum_{i}^{n} |x_i - \bar{x}|$$

Weitsprung: 7 m, 8 m, 10 m

$$\bar{x} = \cancel{12,5} \quad \frac{7+8+10}{3} = \underline{\underline{8,33}}$$

$$d = \frac{1}{3} \cdot \left(|7 - 12,5| + |8 - 12,5| + |10 - 12,5| \right)$$

Bei Häufigkeitsverteilung:

$$d = |x_1 - \bar{x}| \cdot h(x_1) + |x_2 - \bar{x}| \cdot h(x_2) + \ldots + |x_n - \bar{x}| \cdot h(x_n)$$

Weitsprung:

x_i	7 m	8 m	10 m
$H(x_i)$	2	3	5
$h(x_i)$	$\frac{2}{10}$	$\frac{3}{10}$	$\frac{5}{10}$

$$\bar{x} = 8,8$$

$$d = |7 - 8,8| \cdot \frac{2}{10} + |8 - 8,8| \cdot \frac{3}{10} + |10 - 8,8| \cdot \frac{5}{10}$$

„Durchschnittlicher Abstand zum Mittelwert"

Mittlere lineare Abweichung vom Median:

(Mittlere absolute Abweichung)

$$\tilde{d} = \frac{1}{n}\left[\underbrace{|x_1-\tilde{x}| + |x_2-\tilde{x}| + \ldots |x_n-\tilde{x}|}\right]$$

$$= \frac{1}{n}\sum_{i}^{n}|x_i-\tilde{x}|$$

Weitsprung: 7m, 8m, 10m

$\tilde{x} = 8$

$$\tilde{d} = \frac{1}{3}\cdot\left[|7-8| + |8-8| + |10-8|\right]$$

bei Häufigkeitstabelle:

2, 2, 3, 4, 4, 5, 5, 6, 6

$\frac{9 \text{ Werte} + 1}{2} = 5.$ Stelle

2, 2, 3, 4, 4, 5, 5, 6, 6, 7

$\frac{4+5}{2} = \underline{\underline{4,5}}$

$\frac{10 \text{ Werte}}{2} = 5.$ Stelle
6. Stelle

Note	1	2	3	4
Anz.	7	8	4	5

$\frac{24 \text{ Werte}}{2} = 12.$
13.

1, 1, 1, …, 2 …,

Varianz

Mittlere quadratische Abweichung vom Mittelwert:

„Durchschnittliche quadratische Abweichung“

$$\bar{s}^2 = \frac{1}{n} \cdot \left[(x_1-\bar{x})^2 + (x_2-\bar{x})^2 + \dots (x_n-\bar{x})^2 \right]$$
$$= \frac{1}{n} \cdot \sum_i^n (x_i - \bar{x})^2$$

Weitsprung: 7 m, 8 m, 10 m

$\bar{x} = 12,5$

$$\bar{s}^2 = \frac{1}{3} \cdot \left[(7-12,5)^2 + (8-12,5)^2 + (10-12,5)^2 \right]$$

Alternative:

$$\bar{s}^2 = \frac{1}{n} \cdot (x_1^2 + x_2^2 + \dots + x_n^2) - \bar{x}^2$$

bei Häufigkeitsverteilung:

$$\bar{s}^2 = (x_1-\bar{x})^2 \cdot h(1) + (x_2-\bar{x})^2 \cdot (h_2) + \dots (x_n-\bar{x})^2 (h_n)$$

Weitsprung:

x_i	7 m	8 m	10 m
$H(x_i)$	2	3	5
$h(x_i)$	$\frac{2}{10}$	$\frac{3}{10}$	$\frac{5}{10}$

$\bar{x} = 8,8$

$$\bar{s}^2 = (7-8,8)^2 \cdot \frac{2}{10} + (8-8,8)^2 \cdot \frac{3}{10} + (10-8,8)^2 \cdot \frac{5}{10}$$

Alternative:

$$\bar{s}^2 = \left(x_1^2 \cdot h(x_1) + x_2^2 \cdot h(x_2) + \dots x_n^2 \cdot h(x_n) \right) - \bar{x}^2$$

Effizientere Variante

$$s^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}^2 = \frac{1}{n}\sum_{i=1}^{k} n_i a_i^2 - \overline{x}^2$$

In Worten: Bilde das arithmetische Mittel der quadrierten Beobachtungswerte und subtrahiere davon das Quadrat des arithmetischen Mittels. Das wird häufig kompakt geschrieben als: $s^2 = \overline{x^2} - \overline{x}^2$.

Beispiel:

x_i	y_i	$x_i \cdot y_i$	x_i^2	y_i^2
15	10	150	225	100
30	8	240	900	64
60	5	300	3600	25
120	3	360	14400	9
180	1	180	32400	1
240	0.5	120	57600	0.25
$\overline{x} = 107.5$	$\overline{y} = 4.583$	$\sum: 1350$	$\sum: 109125$	$\sum: 199.25$

Somit ergibt sich $s_X^2 = \overline{x^2} - \overline{x}^2 = \frac{1}{6} \cdot 109125 - 107.5^2 = 6631.25$ und $s_Y^2 = \overline{y^2} - \overline{y}^2 = \frac{1}{6} \cdot 199.25 - 4.583^2 \approx 12.20$.

Beispiel: MAD, Varianz, Standardabweichung

mittlere absolute Abweichung

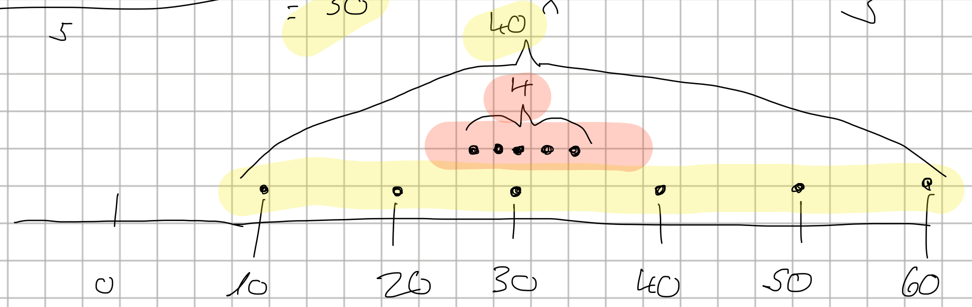
Geg: 10, 20, 30, 40, 50

Geg$_2$: 28, 29, 30, 31, 32

(Arithmetisches Mittel)

$$\bar{x} = \frac{10+20+30+40+50}{5} = 30$$

$$\bar{x} = \frac{28+29+30+31+32}{5} = 30$$



mittlere absolute Abweichung (MAD)

$$= \frac{|10-30| + |20-30| + |30-30| + |40-30| + |50-30|}{5}$$

(durch Betrag $|x|$ nicht negativ)

$$= \frac{20 + 10 + 0 + 10 + 20}{5} = \frac{60}{5} = \underline{\underline{20}}$$

Varianz

$$s^2 = \frac{(10-30)^2 + (20-30)^2 + (30-30)^2 + (40-30)^2 + (50-30)^2}{5}$$

(durch Quadrat x^2 nicht negativ)

$$= \frac{400 + 100 + 0 + 100 + 400}{5} = \frac{1000}{5} = \underline{\underline{200}}$$

Standardabweichung

$$s = \sqrt{s^2} = \sqrt{200} = \underline{\underline{14{,}1}}$$

$$\bar{x} = \frac{2+1+0+1+2}{5} = \frac{6}{5} = \underline{\underline{1{,}2}}$$

$$s^2 = \frac{4+1+0+1+4}{5} = \frac{10}{5} = \underline{\underline{2}}$$

$$s = \sqrt{2} = \underline{\underline{1{,}4}}$$

$$\bar{x} = \underline{\underline{20}}$$

$$s^2 = \underline{\underline{200}}$$

$$s = \underline{\underline{14{,}1}}$$

Die Streuung von ist viel grösser!

Harmonisches Mittel

5 %, 7 %, 7 %, 8 %

$$\bar{x}_h = \frac{4}{\frac{1}{5} + \frac{1}{7} + \frac{1}{7} + \frac{1}{8}} \approx 6{,}549\ldots$$

n <u>positive</u> Zahlen: $x_1, x_2 \ldots x_n$

$$\bar{x}_h = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \ldots \frac{1}{x_n}}$$

Geometrisches Mittel

z.B. Wachstum Kapital

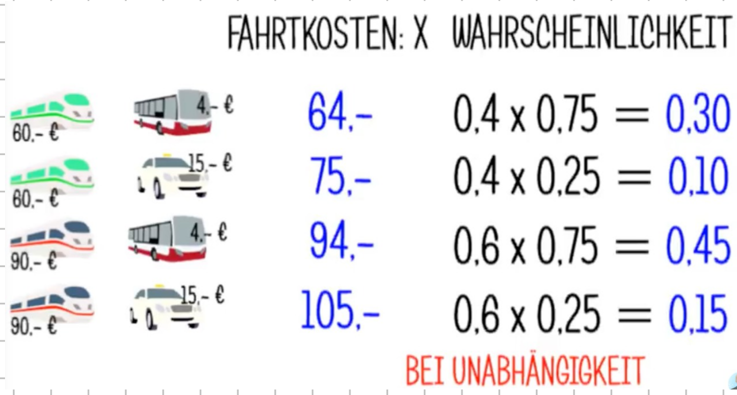
5 %, 7 %, 7 %, 8 %

$$\hat{g} = \sqrt[4]{5 \cdot 7 \cdot 7 \cdot 8} \approx 6{,}65\ldots$$

n <u>positive</u> Zahlen: $g_1, g_2 \ldots g_n$

$$\hat{g} = \sqrt[n]{g_1 \cdot g_2 \cdot \ldots \cdot g_n}$$

Zufallsvariable



X = Eine Grösse die verschiedene Zahlenwertte mit einer gewissen Wahrscheinlichkeit annimmt.

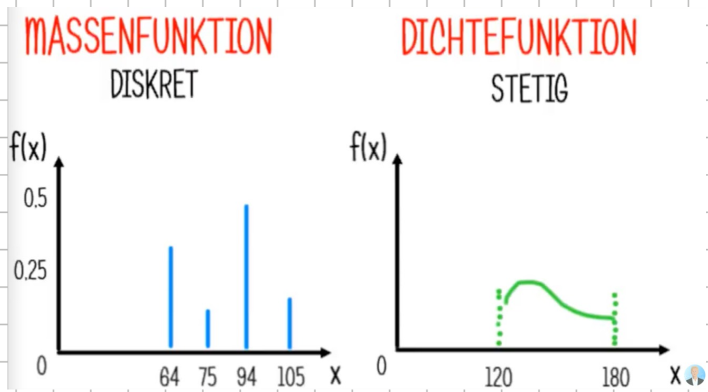
Wahrscheinlichkeitsfunktion

Massenfunktion

$$f(x) = P(X = x) = \begin{cases} 0{,}3 & \text{für } x = 64 \\ 0{,}1 & \text{für } x = 75 \\ 0{,}45 & \text{für } x = 94 \\ 0{,}15 & \text{für } x = 105 \\ 0 & \text{sonst} \end{cases}$$

Alle Wahrscheinlichkeiten zusammengezählt müssen imm 1 ergeben!

Wahrscheinlichkeitsdichte
Dichtefunktion



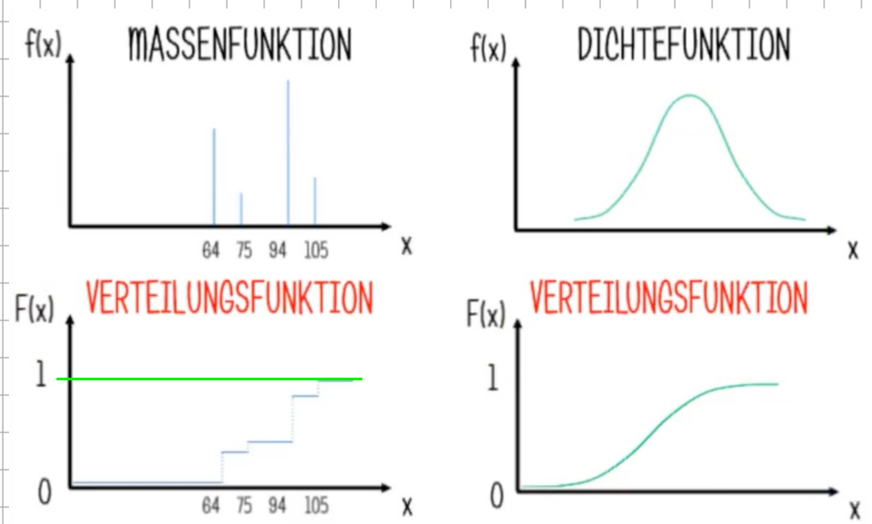
Verteilungsfunktion
(Summenhäufigkeit der Wahrscheinlichkeitsfunktion)

Teilsumme von Einzelwahrscheinlichkeiten

Mit F(x) wird dei Vertilungsfunktion bezeichnet. Diese gibt an, mir welcher Wahrscheinlichkeit die Zufallsvariable höchstens einen Wert x annimmt.

In der Verteilungsfunktion geht es immer um die höchste Wahrscheinlichkeit.

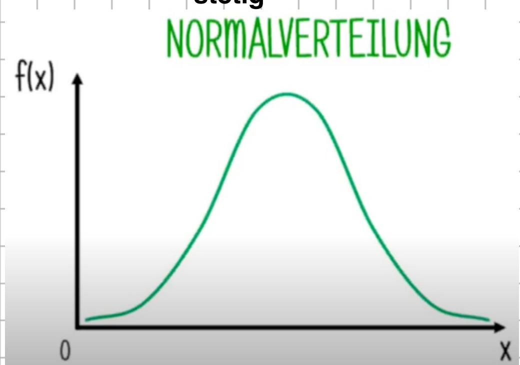
Diskret **Stetig**



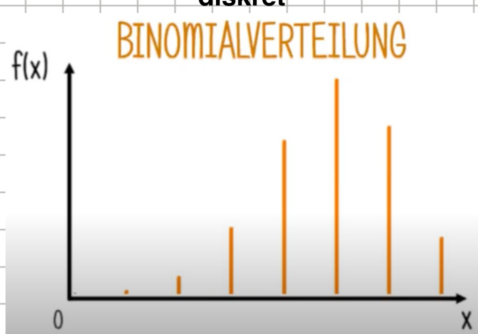
+0.3 +0.1 +0.45 +0.15

Zufallsvariablen:

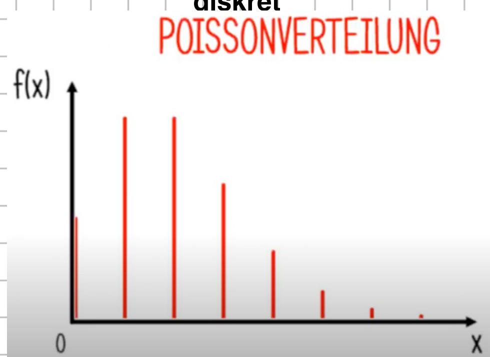
stetig



diskret



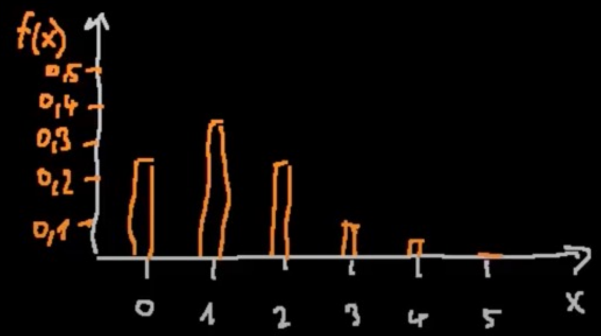
diskret

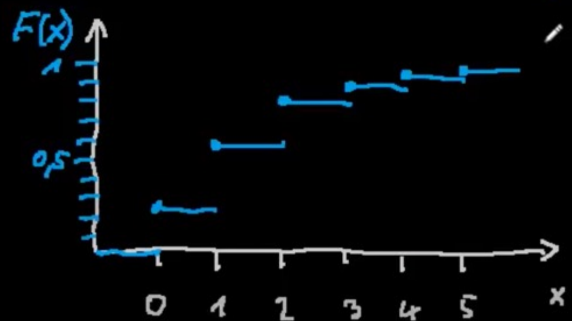


Grundlagen zur Wahrscheinlichkeitsfunktion und Verteilungsfunktion

100 Studenten wurden nach der Anzahl ihrer Geschwister befragt.
Die Befragung ergab folgende Häufigkeitstabelle:

Geschwisteranzahl x_i	0	1	2	3	4	5	Σ
$h(x_i)$	25	35	25	10	4	1	
$f(x)$	0,25	0,35	0,25	0,1	0,04	0,01	1 Wahrscheinlichkeitsf.
$F(x)$	0,25	0,6	0,85	0,95	0,99	1	Verteilungsfunktion





$$f(x) = P(X = x) = \begin{cases} 0{,}25 & \text{für } x = 0 \\ 0{,}35 & \text{für } x = 1 \\ 0{,}25 & \text{für } x = 2 \\ 0{,}1 & \text{für } x = 3 \\ 0{,}04 & \text{für } x = 4 \\ 0{,}01 & \text{für } x = 5 \\ 0 & \text{sonst} \end{cases}$$

$$F(x) = \begin{cases} 0 & \text{für } x < 0 \\ 0{,}25 & \text{für } 0 \le x < 1 \\ 0{,}6 & \text{für } 1 \le x < 2 \\ 0{,}85 & \text{für } 2 \le x < 3 \\ 0{,}95 & \text{für } 3 \le x < 4 \\ 0{,}99 & \text{für } 4 \le x < 5 \\ 1 & \text{für } 5 \le x \end{cases}$$

Zufallsexperiment

Zufallsvariable = X

Erwartungswert = E(X)
Gewichtettes arithmetisches Mittel der Wahrscheinlichkeitsfunktion

$$E(X) = \sum x_i \cdot f(x) = 0 \cdot 0{,}25 + 1 \cdot 0{,}35 + 2 \cdot 0{,}25 + 3 \cdot 0{,}1 + 4 \cdot 0{,}04 + 5 \cdot 0{,}01$$
$$= 0{,}35 + 0{,}5 + 0{,}3 + 0{,}16 + 0{,}05 = 1{,}36$$

Varianz:
Var(x)

$$\tilde{s}^2 = \frac{1}{n}\sum(x_i - \bar{x})^2 \cdot h(x) = \sum(x_i - \bar{x})^2 \cdot f(x) = \sum x_i^2 \cdot f(x) - \bar{x}^2 = \overline{x^2} - \bar{x}^2$$

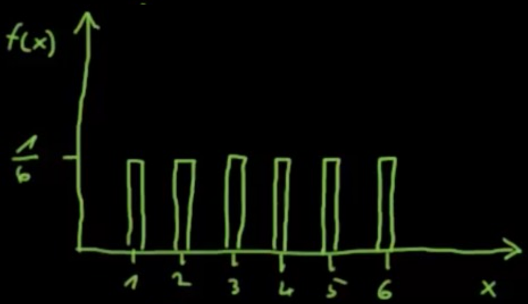
$$\mathrm{Var}(x) = \sum(x_i - E(x))^2 \cdot f(x)$$
$$= (0-1{,}36)^2 \cdot 0{,}25 + (1-1{,}36)^2 \cdot 0{,}35 + (2-1{,}36)^2 \cdot 0{,}25 + (3-1{,}36)^2 \cdot 0{,}1 + (4-1{,}36)^2 \cdot 0{,}04 + (5-1{,}36)^2 \cdot 0{,}01$$
$$= 1{,}36^2 \cdot 0{,}25 + 0{,}36^2 \cdot 0{,}35 + 0{,}64^2 \cdot 0{,}25 + 1{,}64^2 \cdot 0{,}1 + 2{,}64^2 \cdot 0{,}04 + 3{,}64^2 \cdot 0{,}01$$
$$= \underline{1{,}29}$$

Massenfunktion (Wahrscheinlichkeitsfunktion)

X - diskret

$$f(x) = P(X = x) = \begin{cases} 0 & \text{für } x \le 0 \\ \frac{1}{6} & \text{für } 0 < x \le 6 \\ 0 & \text{für } 6 < x \end{cases}$$

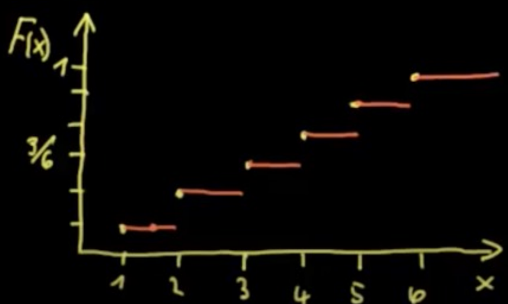
$X =$  $\{1, 2, 3, 4, 5, 6\}$



$$E(x) = \sum_{i=1}^{n} x_i \cdot f(x_i)$$
$$= 1 \cdot \tfrac{1}{6} + 2 \cdot \tfrac{1}{6} + 3 \cdot \tfrac{1}{6} + 4 \cdot \tfrac{1}{6} + 5 \cdot \tfrac{1}{6} + 6 \cdot \tfrac{1}{6}$$
$$= \tfrac{1}{6} + \tfrac{2}{6} + \tfrac{3}{6} + \tfrac{4}{6} + \tfrac{5}{6} + \tfrac{6}{6} = \tfrac{21}{6} = 3{,}5$$

$$\sum_{x=1}^{6} f(x) = 1 \qquad \sum_{i=1}^{n} f(x_i) = 1$$

$$F(x) = P(X \le x) \qquad F(1) = P(X \le 1)$$



Es kann genau für die vorgegebenen Werte eine Wahrscheinlichkeit berechnent werden. Werte dazwischen existieren nicht.

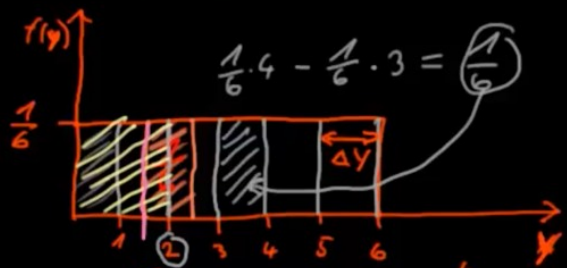
Dichtefunktion (Wahrscheinlichkeitsdichte)

Die Wahrscheinlichkeit von einem bestimmten Bereich.

Y - stetig

$$f(y) = \begin{cases} 0 & \text{für } y \le 0 \\ \frac{1}{6} & \text{für } 0 < y \le 6 \\ 0 & \text{für } 6 < y \end{cases}$$

$Y =$  $\{0 - 6\}$



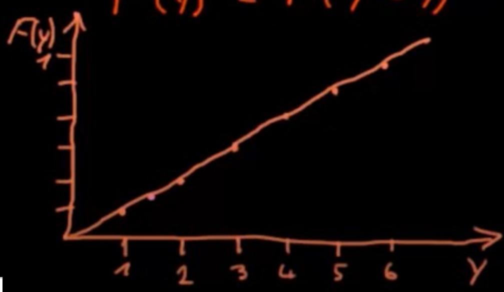
$$\tfrac{1}{6} \cdot 4 - \tfrac{1}{6} \cdot 3 = \tfrac{1}{6}$$

$$E(Y) = \int_{-\infty}^{\infty} y \cdot f(y) \cdot dy = \int_{0}^{6} y \cdot f(y) \cdot dy$$
$$= \int_{0}^{6} \tfrac{1}{6} y \cdot dy = \tfrac{1}{12} y^2 \Big|_0^6$$
$$= \tfrac{1}{12} \cdot 36 - \tfrac{1}{12} \cdot 0 = \tfrac{36}{12} = 3$$

$$\sum_{y=1}^{6} f(y) \cdot \Delta y = 1 \qquad \lim_{n \to \infty} \sum_{i=1}^{n} f(y_i) \cdot \Delta y = 1$$

$$\Rightarrow \int_{-\infty}^{\infty} f(y) \cdot dy$$

$$F(y) = P(Y \le y)$$



Es kann für jeden Wert eine entsprechende Wahrscheinlichkeit bestimmt werden.

$$\tilde{s}^2 = \frac{1}{n}\sum(x_i-\bar{x})^2\cdot h(x) = \sum(x_i-\bar{x})^2\cdot f(x) = \sum x_i^2\cdot f(x) - \bar{x}^2 = \overline{x^2} - \bar{x}^2$$

$$E(x) = \sum_{i=1}^{n} x_i \cdot f(x_i)$$
$$= 1\cdot\tfrac{1}{6} + 2\cdot\tfrac{1}{6} + 3\cdot\tfrac{1}{6} + 4\cdot\tfrac{1}{6} + 5\cdot\tfrac{1}{6} + 6\cdot\tfrac{1}{6}$$
$$= \tfrac{1}{6} + \tfrac{2}{6} + \tfrac{3}{6} + \tfrac{4}{6} + \tfrac{5}{6} + \tfrac{6}{6} = \tfrac{21}{6} = 3{,}5$$

$$Var(x) = \sum_{i=1}^{n} [x_i - E(x)]^2 \cdot f(x)$$
$$= (-2{,}5)^2\cdot\tfrac{1}{6} + (-1{,}5)^2\cdot\tfrac{1}{6} + (-0{,}5)^2\cdot\tfrac{1}{6}$$
$$+ 0{,}5^2\cdot\tfrac{1}{6} + 1{,}5^2\cdot\tfrac{1}{6} + 2{,}5^2\cdot\tfrac{1}{6}$$
$$= \tfrac{6{,}25}{6} + \tfrac{2{,}25}{6} + \tfrac{0{,}25}{6} + \tfrac{0{,}25}{6} + \tfrac{2{,}25}{6} + \tfrac{6{,}25}{6}$$
$$= \tfrac{12{,}5}{6} + \tfrac{5}{6} = \tfrac{17{,}5}{6} = 2{,}92$$

$$E(y) = \int_{-\infty}^{\infty} y\cdot f(y)\cdot dy = \int_{0}^{6} y\cdot f(y)\cdot dy$$
$$= \int_0^6 \tfrac{1}{6} y\cdot dy = \tfrac{1}{12} y^2 \Big|_0^6$$
$$= \tfrac{1}{12}\cdot 36 - \tfrac{1}{12}\cdot 0 = \tfrac{36}{12} = 3$$

$$Var(y) = \int_{-\infty}^{\infty} [y - E(y)]^2\cdot f(y)\cdot dy$$
$$= \int_{-\infty}^{\infty} y^2\cdot f(y) - [E(y)]^2\cdot dy$$
$$= \int_0^6 y^2\cdot f(y) - [E(y)]^2 = \int_0^6 \tfrac{1}{6} y^2 - 9$$
$$= \tfrac{1}{18} y^3 \Big|_0^6 - 9 = \tfrac{1}{18}\cdot 6^3 - \tfrac{1}{18}\cdot 0^3 - 9$$
$$= 12 - 0 - 9 = 3$$

6.4.3 Eigenschaften des Erwartungswerts

Sei X eine ZV mit Erwartungswert μ und g eine Funktion. Dann gilt:

i) $E(X-\mu) = 0$.
ii) $E(g(X)) = \sum_{x_i \in W} g(x_i) f(x_i)$ (falls X diskret) bzw. $E(g(X)) = \int_{-\infty}^{\infty} g(x) f(x)\, dx$ (falls X stetig) .
iii) $E(a) = a$, d.h. der EW einer Konstanten a ist gleich dieser Konstanten.
iv) $E(b\cdot g(X)) = b\cdot E(g(X))$ für jede Konstante b, insbesondere ist $E(bX) = bE(X)$.
v) $E(g_1(X) + g_2(X)) = E(g_1(X)) + E(g_2(X))$, insbesondere ist $E(X_1 + X_2) = E(X_1) + E(X_2)$.
vi) $E(X+a) = E(X) + a$.

6.5.3 Eigenschaften der Varianz

Auch für die Varianz gelten gewisse Rechenregeln, die ich hier kurz zusammenfassen will: Sei X eine ZV mit Erwartungswert μ und b eine reelle Zahl. Dann gilt:

i) $var(a) = 0$ (Eine konstante ZV streut nicht!) .
ii) $var(X+a) = var X$ (Wird eine ZV um einen konstanten Betrag verschoben, ändert sich die Varianz nicht.) .
iii) $var(bX) = b^2 var X$ bzw. $\sigma_{bX} = |b|\sigma_X$ (Ein konstanter Faktor wird also mit dem Quadrat aus der Varianz "herausgezogen"!) .
iv) $var X = E(X^2) - \mu^2$.
v) $var X = E(X-b)^2 - (\mu - b)^2$ (Das ist der sogenannte Steinersche Verschiebungssatz. 4. ist davon ein Spezialfall ($b = 0$).) .
vi) $var(bX + a) = b^2 var X$ (Das ist eine Zusammenfassung von Eigenschaften 2. und 3.) .

Achtung: Es gilt also i.A. nicht $var(X_1 + X_2) = var(X_1) + var(X_2)$!

Wann welche Verteilung benützen?

	DISKRETE SCHRITTE?	STETIGE SPANNE?
ANZAHL?	BINOMIAL ~~HYPERGEOMETRISCH~~	POISSON
ERSTES MAL?	GEOMETRISCH	EXPONENTIAL

Nicht Modulrelevant

~~GESAMT N=10~~

~~DEFEKT m=2~~

~~VERBAUT n=5~~

~~GESUCHT: WAHRSCHEINLICHKEIT FÜR x=1~~

~~HYPERGEOMETRISCH~~ $\cancel{f_h(x|N;m;n) = \frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}}}$

GESAMT ...

DEFEKT $\Theta=0{,}2$

ERHALTEN $n=10$

GESUCHT: WAHRSCHEINLICHKEIT FÜR $x=2$

BINOMIAL $f_b(x|\Theta;n) = \binom{n}{x} \cdot \Theta^x \cdot (1-\Theta)^{n-x}$

Binomialverteilung wenn die Grundgesamtheit sehr gross ist und die Entnommenen Stücke die Grundgesamtheit nicht beeinflussen.

GEOMETRISCH

$$f_g(x|\Theta) = (1-\Theta)^{x-1} \cdot \Theta$$

$\Theta=0{,}2$

GESUCHT: WAHRSCHEINLICHKEIT FÜR $x=4$

EXPONENTIAL

$$F_e(x|\lambda) = 1 - e^{-\lambda x}$$

$\lambda=1$

$x=1$

POISSON

$$f_p(x|\lambda) = \frac{\lambda^x}{x!} \cdot e^{-\lambda}$$

$\lambda=3$

$x=2$

Gefragt ist	Bezugsgröße ist diskret (z.B. Schrittzahl)	stetig (z.B. Zeitspanne)
nach der Häufigkeit, mit der das interessierende Merkmal eintritt	nur Anteilswert Θ ist bekannt: Binomialverteilung Zahlen zur Gesamtheit (N und M) sind bekannt: hypergeometrische Verteilung	Poisson-Verteilung
wann das interessierende Ereignis eintritt	geometrische Verteilung	Exponentialverteilung

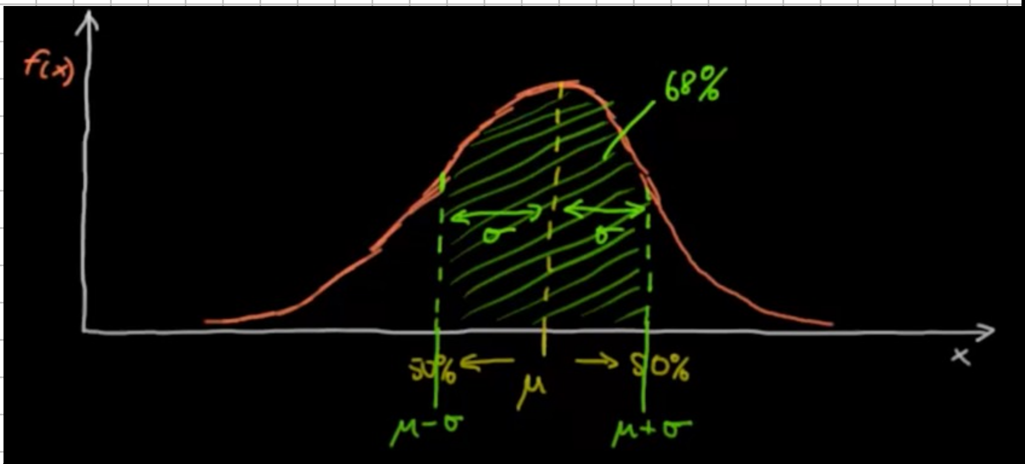
Verteilungen

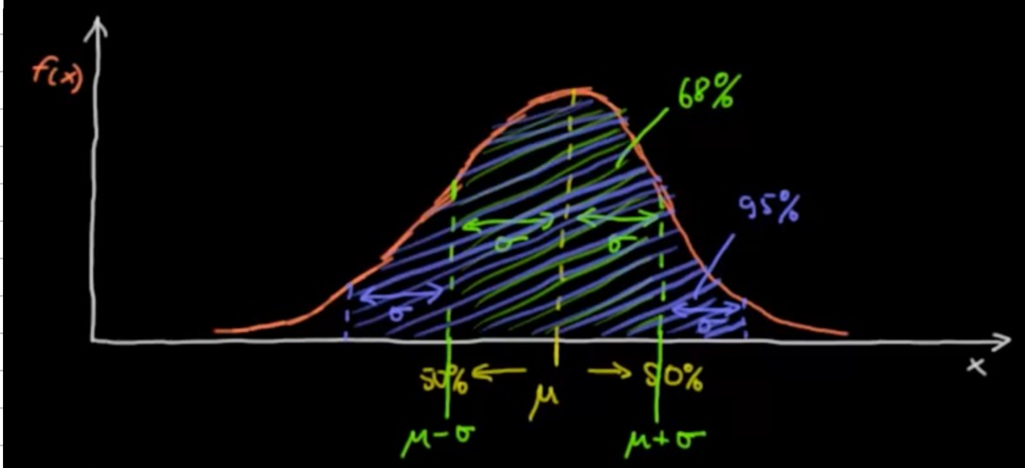
Normalverteilung

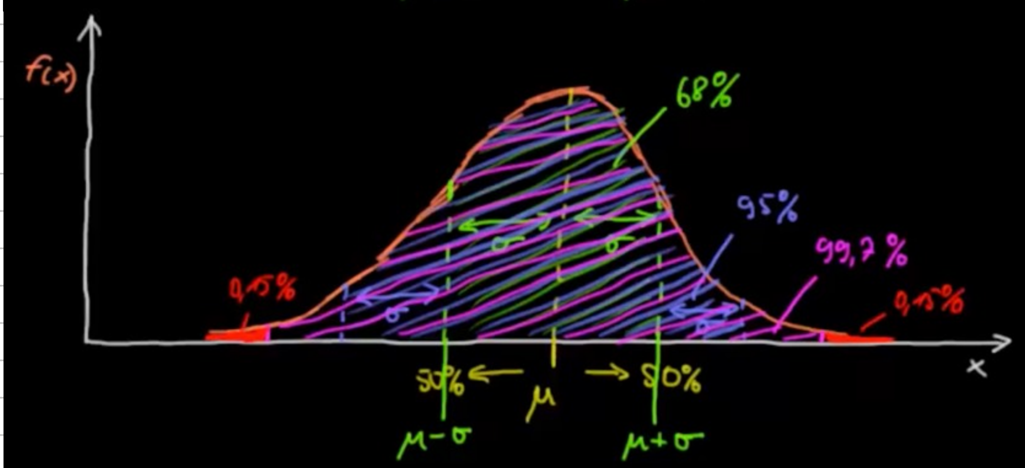
μ "müh" -> Miittelwert

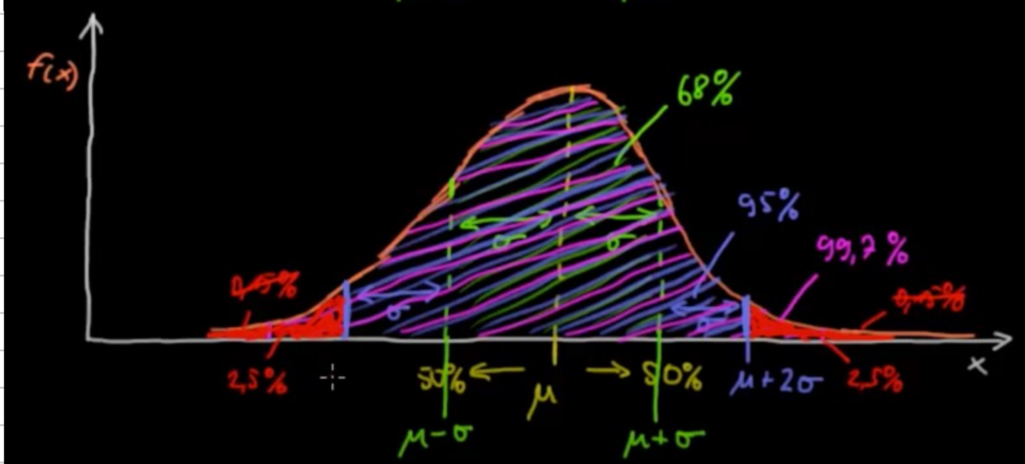
σ "sigma" -> Standardabweichung

Varianz σ^2
standardabw. σ
68 – 95 – 99,7
1 – 2 – 3 σ









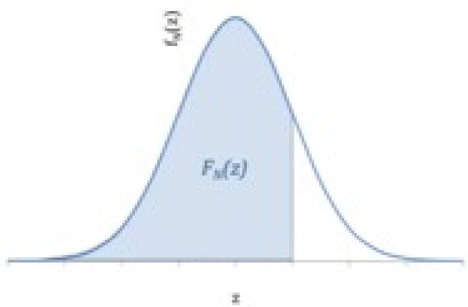
Beispiel: 5/4 = 1.25

σ

Anhang A: Tabelle der Standardnormalverteilung

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990

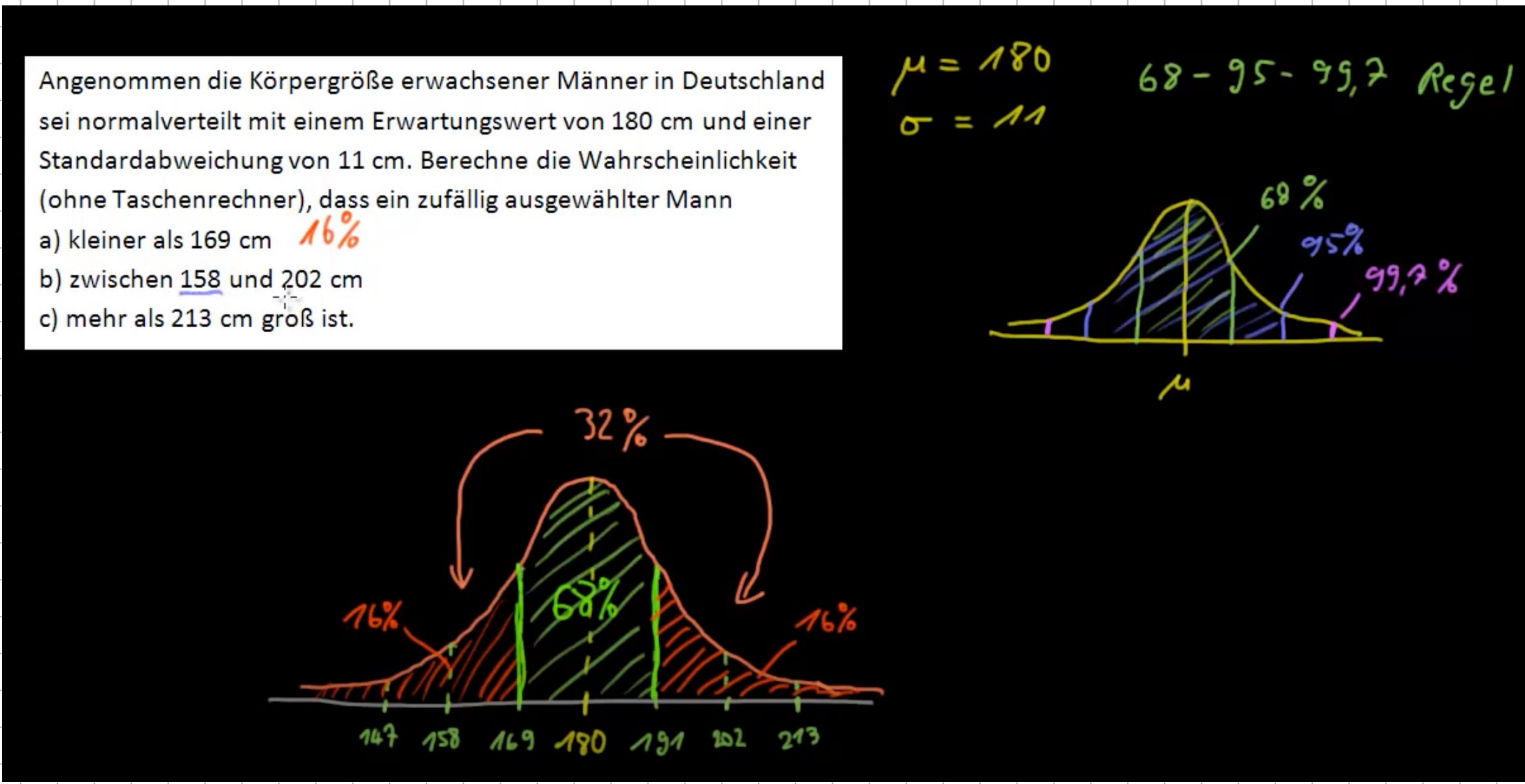
Tabelle 78: Tabelle der Standardnormalverteilung

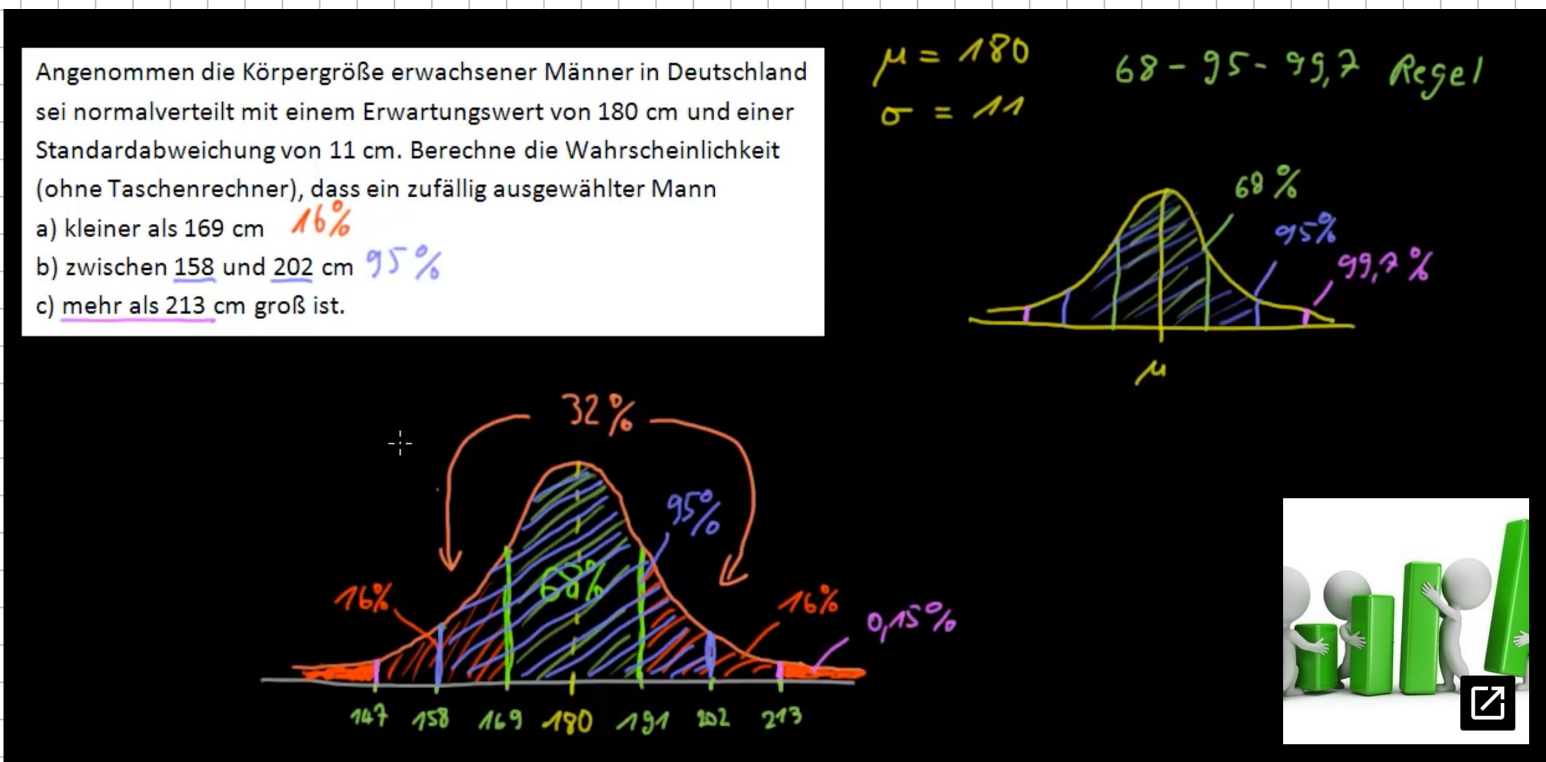


- Tabelliert sind Werte der Verteilungsfunktion $F_N(z)$
- Wahrscheinlichkeitsmassen für Werte $z < 0$ berechnen sich nach $F_N(-z) = 1 - F_N(z)$
- Beispiele:

$$F_N(1{,}57) = 0{,}9418$$

$$F_N(-1{,}57) = 1 - F_N(1{,}57) = 0{,}0582$$


Angenommen die Körpergröße erwachsener Männer in Deutschland sei normalverteilt mit einem Erwartungswert von 180 cm und einer Standardabweichung von 11 cm. Berechne die Wahrscheinlichkeit (ohne Taschenrechner), dass ein zufällig ausgewählter Mann
a) kleiner als 169 cm 16%
b) zwischen 158 und 202 cm
c) mehr als 213 cm groß ist.
μ = 180
σ = 11
68 - 95 - 99,7 Regel
68 %
95%
99,7 %
μ
32 %
16%
68%
16%
147 158 169 180 191 202 213


Angenommen die Körpergröße erwachsener Männer in Deutschland sei normalverteilt mit einem Erwartungswert von 180 cm und einer Standardabweichung von 11 cm. Berechne die Wahrscheinlichkeit (ohne Taschenrechner), dass ein zufällig ausgewählter Mann
a) kleiner als 169 cm 16%
b) zwischen 158 und 202 cm 95 %
c) mehr als 213 cm groß ist.
μ = 180
σ = 11
68 - 95 - 99,7 Regel
68 %
95%
99,7 %
μ
32 %
95%
16%
68%
16%
0,15 %
147 158 169 180 191 202 213

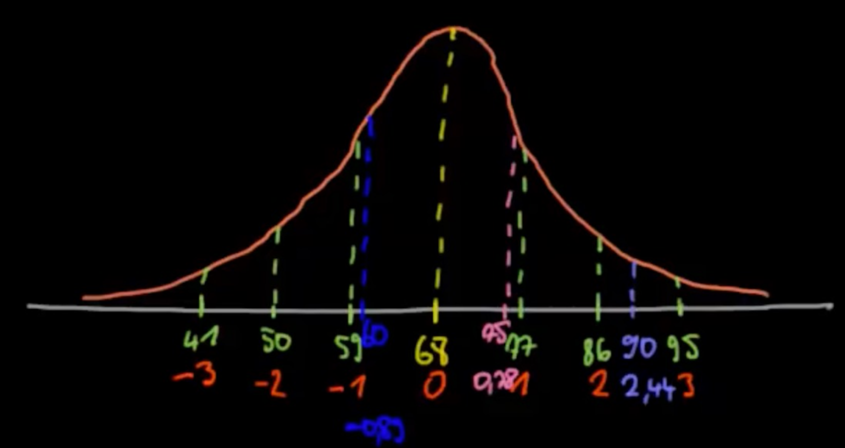
Die Ergebnisse von Statistik-Klausuren an der FernUni Hagen seien annähernd normalverteilt mit einer mittleren Punktzahl von μ = 68 Prozentpunkten und einer Standardabweichung von σ = 9 Prozentpunkten. Berechne die Z-Werte für die folgenden Punktzahlen und ermittle die Wahrscheinlichkeiten, dass ein zufällig ausgewählter Student mindestens die jeweilige Note erreicht.

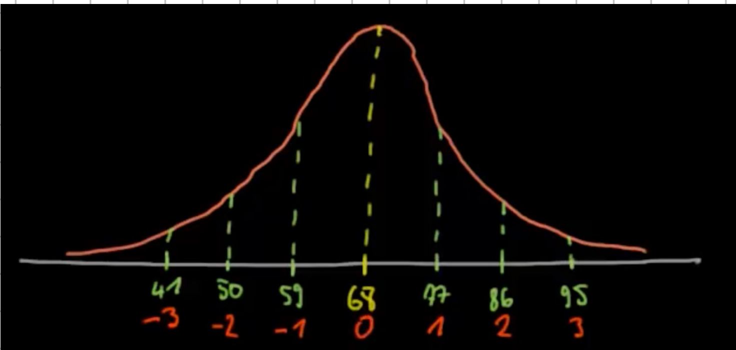
90 Punkte – 1 $\quad \frac{90-68}{9} = \frac{22}{9} = 2{,}44$

75 Punkte – 2 $\quad \frac{75-68}{9} = \frac{7}{9} = 0{,}78$

60 Punkte – 3 $\quad \frac{60-68}{9} = -0{,}89$

50 Punkte – 4



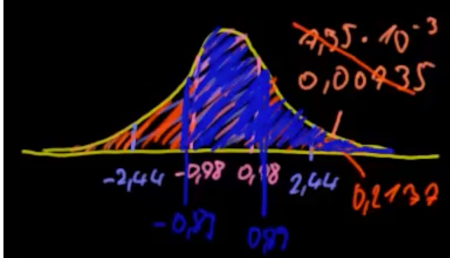


annähernd normalverteilt
Standardabweichung
Punktzahlen und ermittle
indestens die jeweilige

75 Punkte – 2 $\quad \frac{75-68}{9} = \frac{7}{9} = 0{,}78 \Rightarrow 21{,}77\,\%$

60 Punkte – 3 $\quad \frac{60-68}{9} = -0{,}89 \Rightarrow 81{,}325\,\%$

50 Punkte – 4 $\quad -2$



$F_z(2{,}44) = 0{,}9853$

$1 - 0{,}9853 = 0{,}0147$

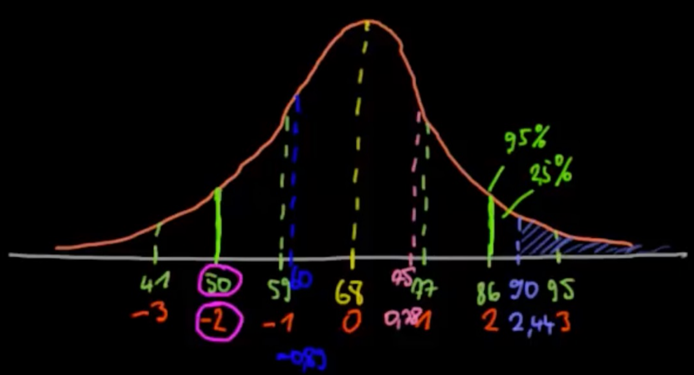
$F_z(0{,}78) = 0{,}5646$

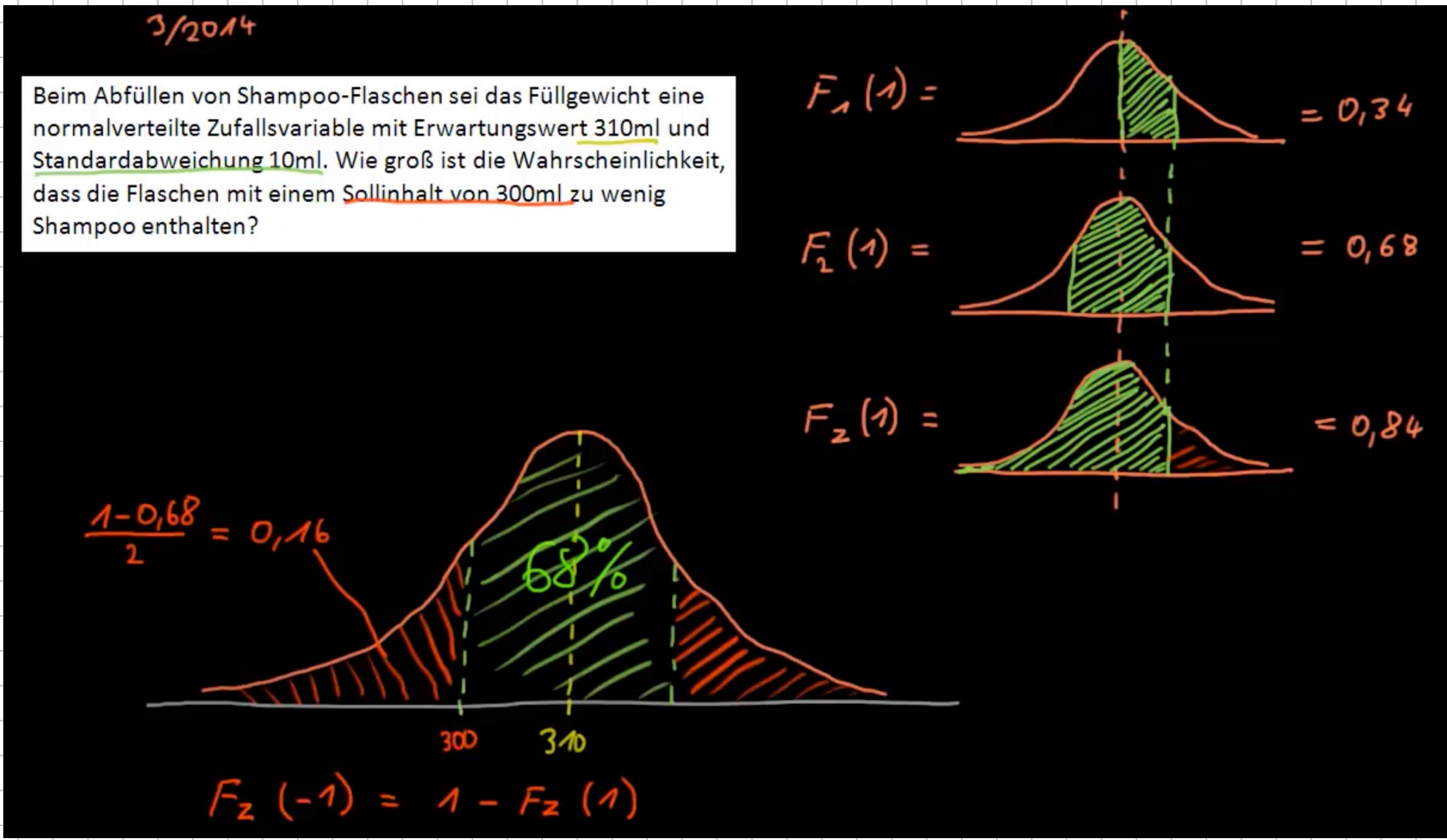
$1 - 0{,}5646 = 0{,}4354$

$F_z(0{,}89) = 0{,}6265$

$1 - 0{,}6265 = 0{,}3735$

$0{,}6265 + 0{,}18675 = 0{,}81325$




3/2014
Beim Abfüllen von Shampoo-Flaschen sei das Füllgewicht eine normalverteilte Zufallsvariable mit Erwartungswert 310ml und Standardabweichung 10ml. Wie groß ist die Wahrscheinlichkeit, dass die Flaschen mit einem Sollinhalt von 300ml zu wenig Shampoo enthalten?
$F_1(1) = \; = 0,34$
$F_2(1) = \; = 0,68$
$F_Z(1) = \; = 0,84$
$\frac{1-0,68}{2} = 0,16$
68%
300
310
$F_Z(-1) = 1 - F_Z(1)$

Poisson - Verteilung

Bsp: Telefonzentrale

im Durchschnitt 7 Anrufe pro Stunde

Wahrsch. für 4 Anrufe in einer Stunde?

Idee: Stunde aufteilen in 60 Minuten

$$X \sim B_{60,p}$$

$$E(X) = 60p = 7 \qquad p = 7/60$$

$$X_n \sim B_{n,7/n}$$

$$P(X_n = 4) = \binom{n}{4} \cdot \left(7/n\right)^4 \cdot \left(1-\frac{7}{n}\right)^{n-4}$$

$$= \frac{n(n-1)(n-2)(n-3)}{1\cdot 2\cdot 3\cdot 4} \cdot \frac{7^4}{n^4} \cdot \left(1-\frac{7}{n}\right)^{-4} \left(1-\frac{7}{n}\right)^n$$

$$= \frac{7^4}{4!} \cdot \frac{n^4-6n^3+11n^2-6n}{n^4} \cdot \left(1-\frac{7}{n}\right)^{-4} \cdot \left(1-\frac{7}{n}\right)^n$$

$$= \frac{7^4}{4!} \cdot \frac{n^4-6n^3+11n^2-6n}{n^4} \cdot \left(1-\frac{7}{n}\right)^{-4} \cdot \boxed{\left(1-\frac{7}{n}\right)^n}$$

$$\Big\downarrow {\scriptstyle n\to\infty}$$

$$\frac{7^4}{4!} \cdot 1 \cdot 1 \cdot \boxed{e^{-7}}$$

$$= \frac{7^4}{4!} \cdot e^{-7}$$

$$P(X=k) = \frac{\lambda^k}{k!} \cdot e^{-\lambda}$$

X ist Poisson-verteilt mit Parameter λ

$$\boxed{E(X) = \lambda}$$

$$Var(X_n) = np(1-p) = E(X_n)\cdot(1-p) \xrightarrow{n\to\infty} \lambda \cdot 1 = \lambda$$

$$P(x=k) = B(n;p;k) \approx \frac{\mu^k}{k!} \cdot e^{-\mu}$$

- große Werte für n, kleine Werte für k

1000 Schrauben, durchsch. sind 2 kaputt

WK, dass genau 5 kaputt sind?

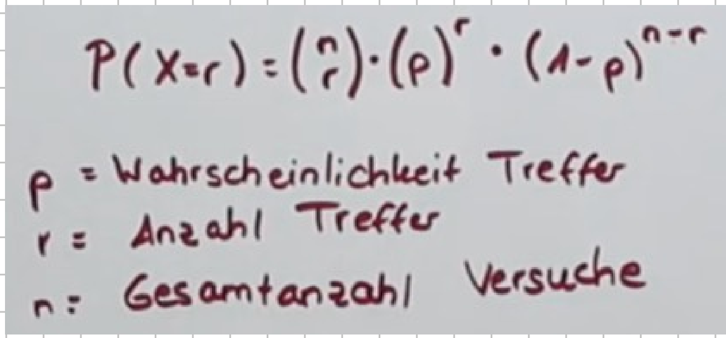
$$P(x=5) = \frac{2^5}{5!} \cdot e^{-2} \approx 0{,}036 \qquad 3{,}6\ \%$$

WK, dass höchstens 5 kaputt sind?

$$P(x \le 5) = P(x=0) + \ldots\ldots + P(x=5)$$

$$\approx \frac{2^0}{0!} \cdot e^{-2} + \ldots\ldots + \frac{2^5}{5!} \cdot e^{-2}$$

Binomialverteilung



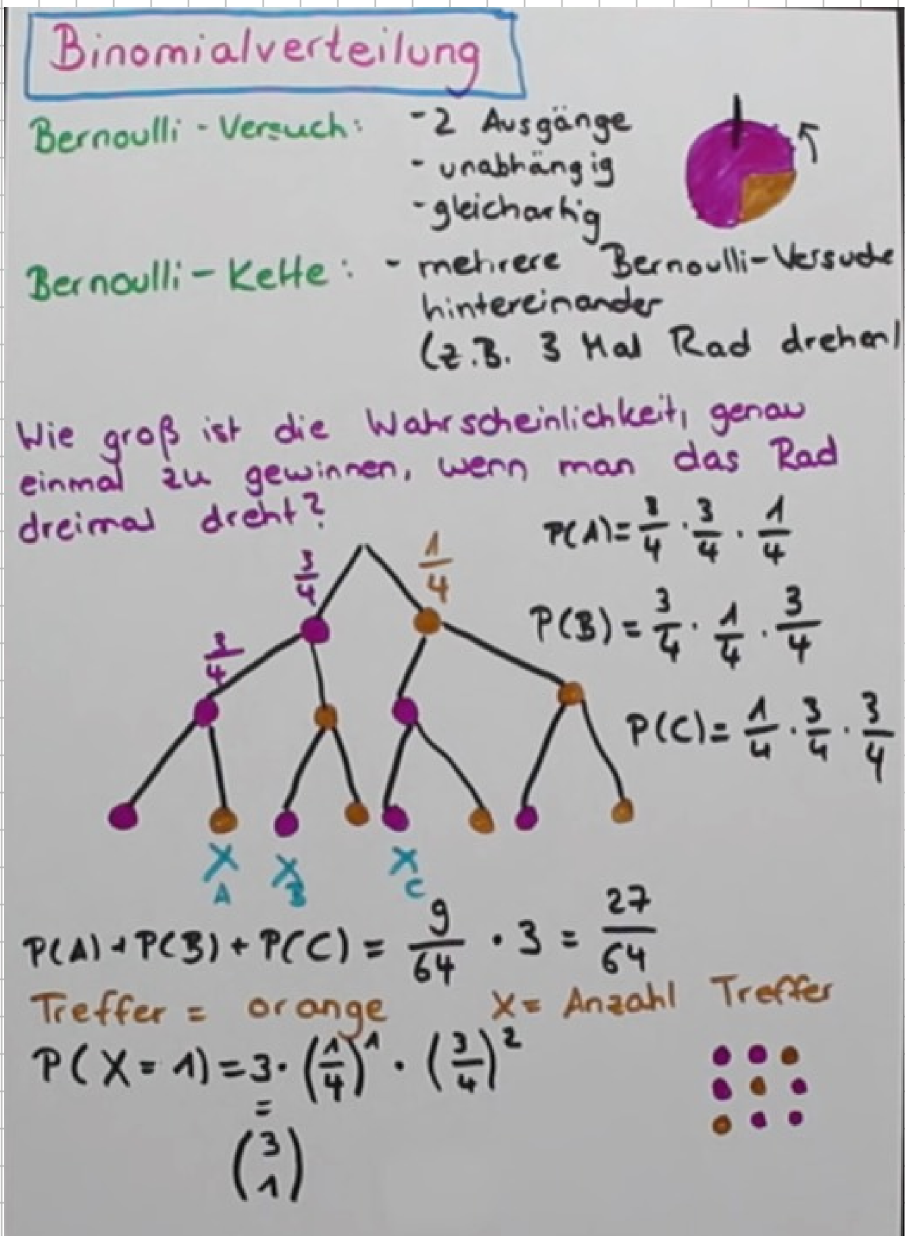
$$P(X=r) = \binom{n}{r} \cdot (p)^r \cdot (1-p)^{n-r}$$

p = Wahrscheinlichkeit Treffer
r = Anzahl Treffer
n = Gesamtanzahl Versuche

- 2 Ausgänge -> o oder 1
- unabhängig
- gleichartig

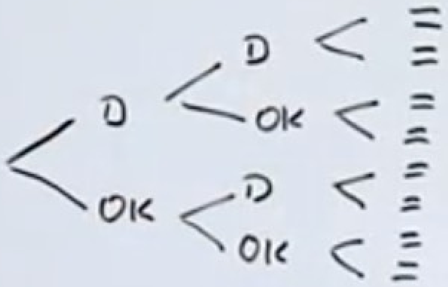
Bernoulli - Versuch

Bernoulli - Kette : Mehrere Bernoulli - Versuche hintereinander



Binominalverteilung (Bernoulli-Kette)

$\boxed{p = 10\,\%}$ aller Schrauben sind defekt.
$\boxed{n = 100}$ Schrauben werden der Tagesproduktion entnommen. Mit welcher WK sind $\boxed{\frac{18}{k}}$ defekt?



$$\binom{100}{18} \cdot 0{,}1^{18} \cdot 0{,}9^{82}$$

$$P(x = \boxed{k}) = \binom{\boxed{n}}{\boxed{k}} \cdot \boxed{p}^{\boxed{k}} \cdot (1-\boxed{p})^{\boxed{n}-\boxed{k}}$$

$$P(x = k) = \binom{n}{k} \cdot p^{k} \cdot (1-p)^{n-k}$$

$$P(x = 18) = \binom{100}{18} \cdot 0{,}1^{18} \cdot 0{,}9^{82}$$

Name	Anwendung	Massen- bzw. Dichtefunktion und Verteilungsfunktion; Erwartungswert und Varianz	Beispielhafte Problemstellung
Binomialverteilung	Vorgang läuft von Schritt zu Schritt unabhängig ab, die Wahrscheinlichkeit für das Auftreten des interessierenden Ereignisses beträgt in jedem Schritt Θ. Gesucht sind Wahrscheinlichkeiten für eine bestimmte Anzahl x von Beobachtungen des interessierenden Merkmals bei n Schritten.	$f_b(x\|\Theta;n) = \binom{n}{x} \cdot \Theta^x \cdot (1-\Theta)^{n-x}$ $F_b(x\|\Theta;n) = \sum_{k=0}^{x} \binom{n}{k} \cdot \Theta^k \cdot (1-\Theta)^{n-k}$ $EX_b = n \cdot \Theta; varX_b = n \cdot \Theta \cdot (1-\Theta)$	Eine Maschine produziert Ausschuss mit Wahrscheinlichkeit $\Theta = 0{,}02$ je Stück. Wie groß ist die Wahrscheinlichkeit, dass in einer Produktion von $n = 1.000$ Stück mehr als 50 ($x \geq 50$) Ausschussstücke enthalten sind?
Geometrische Verteilung	Wie Binomialverteilung; gesucht sind Wahrscheinlichkeiten für die erstmalige Beobachtung des interessierenden Ereignisses in einem bestimmten Schritt x.	$f_g(x\|\Theta) = (1-\Theta)^{x-1} \cdot \Theta$ $F_g(x\|\Theta) = 1 - (1-\Theta)^x$ $EX_g = \frac{1}{\Theta}; varX_g = \frac{1-\Theta}{\Theta^2}$	Die Wahrscheinlichkeit, dass ein Fahrstuhl bei einer Benutzung versagt, liegt bei 0,05% ($\Theta = 0{,}0005$). Wie groß ist die Wahrscheinlichkeit, dass der Fahrstuhl frühestens bei der 1.000sten ($x \geq 1000$) Benutzung versagt?
Hypergeometrische Verteilung	Es liegt eine endliche Gesamtheit mit N Elementen vor, wovon M-viele die interessierende Merkmalsausprägung tragen. Es werden n Elemente zufällig ohne Zurücklegen entnommen. Gesucht sind Wahrscheinlichkeiten für eine bestimmte Anzahl x von Elementen mit der Merkmalsausprägung in der Stichprobe.	$f_h(x\|N;M;n) = \frac{\binom{M}{x} \cdot \binom{N-M}{n-x}}{\binom{N}{n}}$ $F_h(x\|N;M;n) = \sum_{k=0}^{x} \frac{\binom{M}{k} \cdot \binom{N-M}{n-k}}{\binom{N}{n}}$ $EX_h = n \cdot \frac{M}{N}$ $varX_h = n \cdot \frac{M}{N} \cdot \left(1 - \frac{M}{N}\right) \cdot \frac{N-n}{N-1}$	In einem Betrieb mit N = 100 Mitarbeitern sind M = 15 unzufrieden mit der Unternehmenskommunikation. Wie groß ist die Wahrscheinlichkeit, dass bei einer Befragung von $n = 20$ zufällig ausgewählten Mitarbeitern kein unzufriedener befragt wird ($x = 0$)?

Name	Anwendung	Massen- bzw. Dichtefunktion und Verteilungsfunktion; Erwartungswert und Varianz	Beispielhafte Problemstellung
Poisson-Verteilung	Ein interessierendes Ereignis tritt erfahrungsgemäß innerhalb einer zeitlichen, räumlichen oder anderweitig stetigen Spanne im Schnitt mit der Häufigkeit λ auf. Gesucht ist die Wahrscheinlichkeit für ein x-maliges Auftreten dieses Ereignisses innerhalb dieser Spanne. Grenzverteilung zur Binomialverteilung bei großem n ($n \geq 100$) und kleinem Θ ($\Theta \leq 0{,}1$).	$f_p(x\|\lambda) = \frac{\lambda^x}{x!} \cdot e^{-\lambda}$ $F_p(x\|\lambda) = e^{-\lambda} \cdot \sum_{k=0}^{x} \frac{\lambda^k}{k!}$ $EX_p = \lambda; varX_p = \lambda$	An einer Tankstelle tanken pro Nacht im Durchschnitt $\lambda = 7$ Fahrzeuge. Wie groß ist die Wahrscheinlichkeit, dass in einer Nacht gar kein Fahrzeug tanken wird ($x = 0$)?
Exponentialverteilung	Verstetigung der geometrischen Verteilung; gesucht sind Wahrscheinlichkeiten für die erstmalige Beobachtung des interessierenden Ereignisses nach Ablauf einer Zeit-, Weg- oder anderen (quasi-) stetigen Spanne. Den Parameter λ schätzt man als Kehrwert der durchschnittlich beobachteten Spanne bis zum Eintritt.	$f_e(x\|\lambda) = \lambda e^{-\lambda x}$ $F_e(x\|\lambda) = 1 - e^{-\lambda x}$ $EX_e = \frac{1}{\lambda}; varX_e = \frac{1}{\lambda^2}$	Die durchschnittliche Brenndauer einer Energiesparlampe beträgt 5.000 Stunden. Wie groß ist die Wahrscheinlichkeit, dass die Lampe maximal 1.000 Stunden ($x \leq 1.000$) brennt?

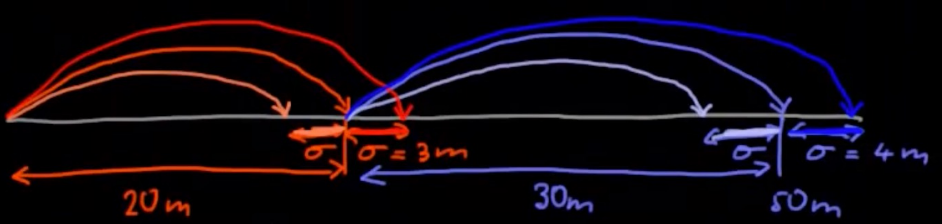
Approximation

Erwartungswert, Varianz, Standardabweichung

Die 10. Klasse der Carl-Friedrich-Gauß-Gesamtschule tritt bei den Bundesjugendspielen an.

Die Ergebnisse beim Werfen seien jeweils für Mädchen und Jungen normalverteilt. Die Mädchen dieser Jahrgangsstufe kommen dabei auf eine durchschnittliche Weite von 20m mit einer Standardabweichung von 3m. Und die Jungen kommen auf eine durchschnittliche Weite von 30m mit einer Standardabweichung von 4m.

Wie groß ist die Wahrscheinlichkeit dafür, dass ein zufällig ausgewähltes Mädchen und ein zufällig ausgewählter Junge gemeinsam in Summe eine Weite von mehr als 60m erreichen?



$W = X + Y$

$E(W) = 50$

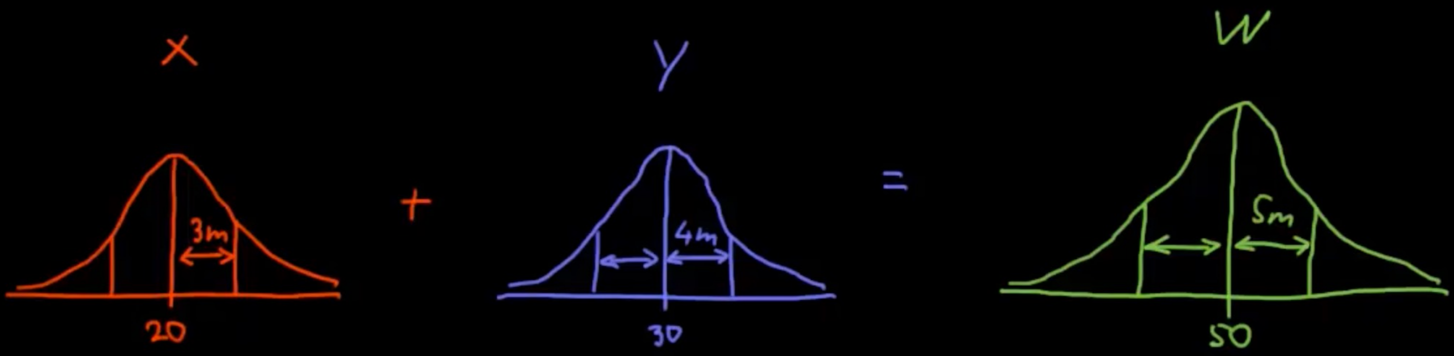
$\tilde{S}_W = 5$

Var(x) = 3^2 = 9

Var(y) = 4^2 = 16

$$E(X+Y) = E(X) + E(Y) = 20 + 30 = 50$$

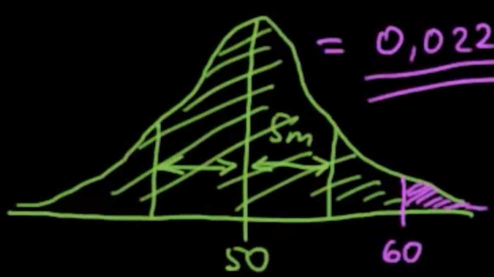
$$Var(X+Y) = Var(X) + Var(Y) = 9 + 16 = 25 \Rightarrow \tilde{S}_{X+Y} = \sqrt{25} = 5$$



W $\quad 1 - F_z(2)$

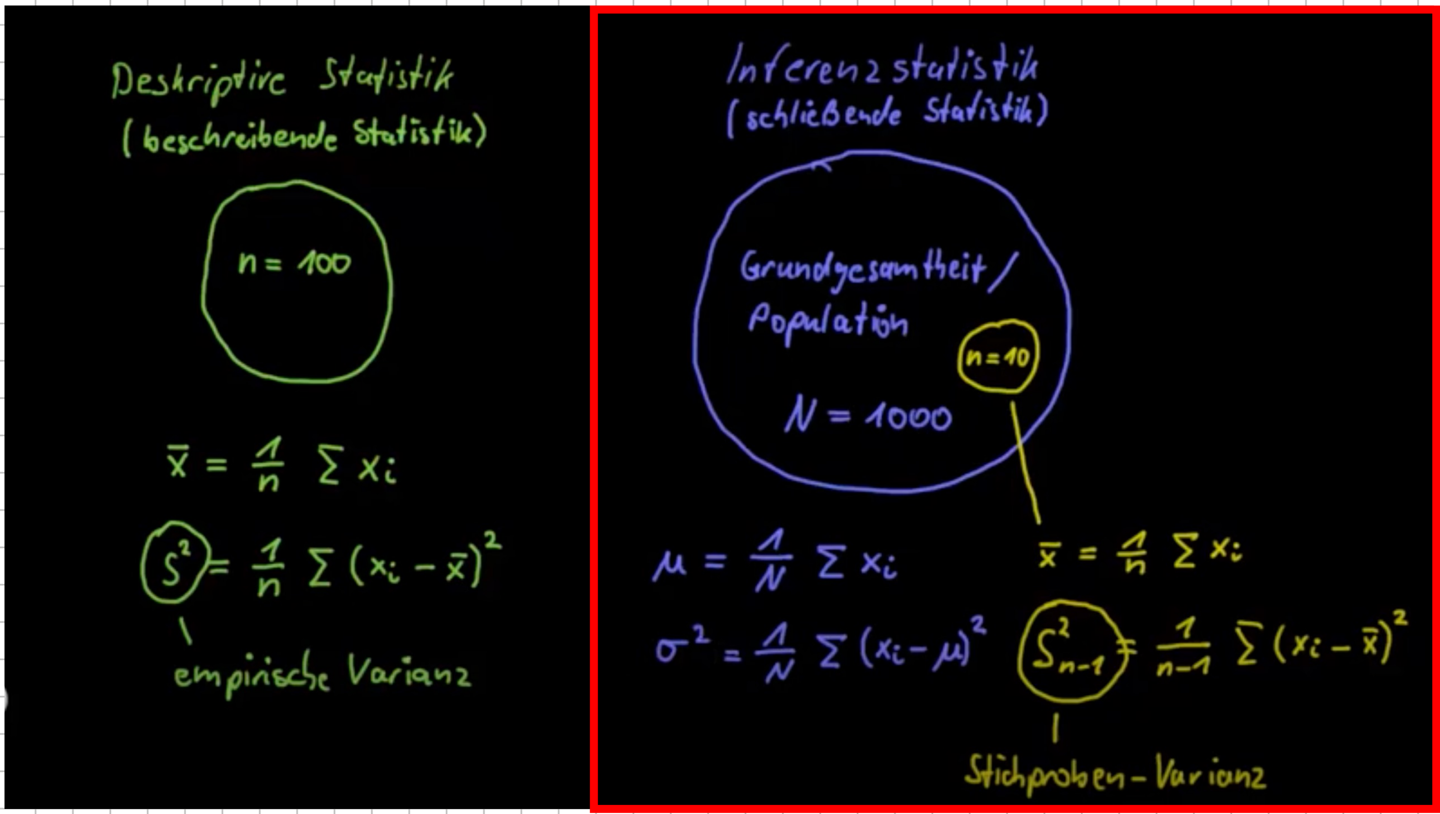
$= 1 - 0{,}9772$

$= \underline{0{,}0228}$



Inferenzstatistik - Schliessende Statistik

Anhand einer Stichprobe werden Rückschlüsse auf eine Grundgesamtheit gezogen.



μ ist der mittelwert der Grundgesamtheit

σ^2 Varianz der Grundgesamtheit

$\bar{x}$ ist der Mittelwert der Stichprobe -> Je grösser desto besser

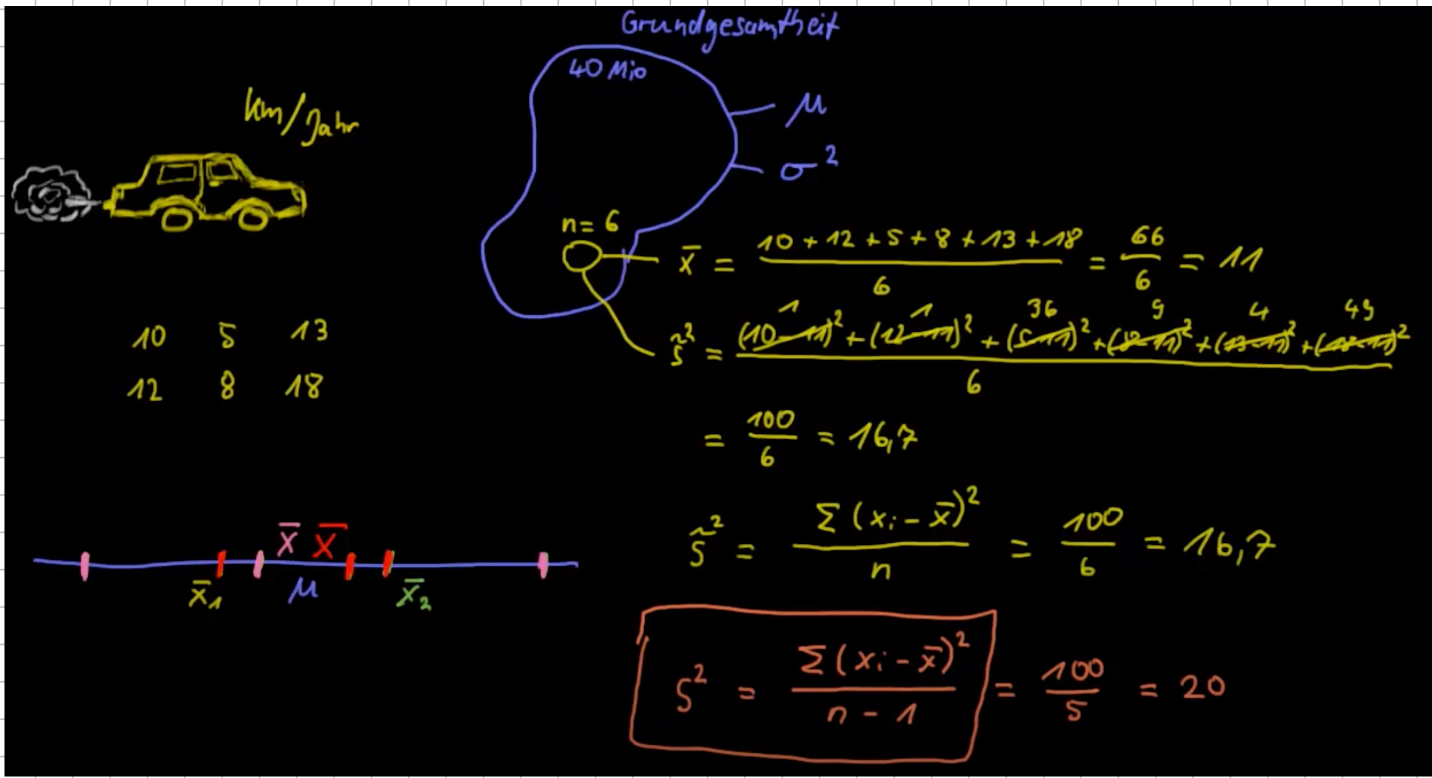
$\tilde{s}^2$ ist die Varianz der Stichprobe “Die durchschnittliche quadratische Abweichung”

s^2 ist die Annahme der Varianz der Grundgesamtheit -> Es wird von ausgegangen, dass die Grundgesamtheit immer eine grössere Streuung hat als die Stichprobe

n ist die Anzahl Stichproben

- Parameter der Grundgesamtheit werden mit griechischen Buchstaben bezeichnet:
  - μ für ein arithmetisches Mittel der Grundgesamtheit,
  - Θ für einen Anteilswert,
  - σ für die Standardabweichung und σ^2 für die Varianz.
  - Die Anzahl der Träger einer bestimmten Eigenschaft bekommen allerdings den Buchstaben M; die Anzahl aller Elemente wird mit N bezeichnet.
- Aus der Stichprobe ermittelte Werte bekommen lateinische Buchstaben:
  - $\bar{x}$ verwendet man für das arithmetische Mittel,
  - p für einen Anteilswert,
  - s für die Standardabweichung und s^2 für die Varianz (bzw. s^* und s^{*2}; siehe G.1.2.4).
  - Die Anzahl der Träger einer bestimmten Eigenschaft in der Stichprobe bekommen den Buchstaben m; der gesamte Umfang der Stichprobe wird mit n bezeichnet.

Inferenzstatistik



μ ist der mittelwert der Grundgesamtheit

σ^2 Varianz der Grundgesamtheit

$\bar{x}$ ist der Mittelwert der Stichprobe -> Je grösser desto besser

$\tilde{s}^2$ s^* ist die empirische Varianz der Stichprobe —> Die durchschnittliche quadratische Abweichung

s^2 s^{*2} ist die Stichprobenvarianz —> Annahme der Varianz der Grundgesamtheit -> Es wird von ausgegangen, dass die Grundgesamtheit immer eine grössere Streuung hat als die Stichprobe

n ist die Anzahl Stichproben

$\hat{\sigma}$ = s ist der Schätzwert für die Standardabweichung

σ ist die Standardabweichung

$\mu \leftarrow \bar{x}$

$\sigma \leftarrow s$

$$\tilde{s}^2 = \frac{\sum(x_i-\bar{x})^2}{n} = \frac{100}{6} = 16{,}7$$

$$s^2 = \frac{\sum(x_i-\bar{x})^2}{n-1} = \frac{100}{5} = 20$$

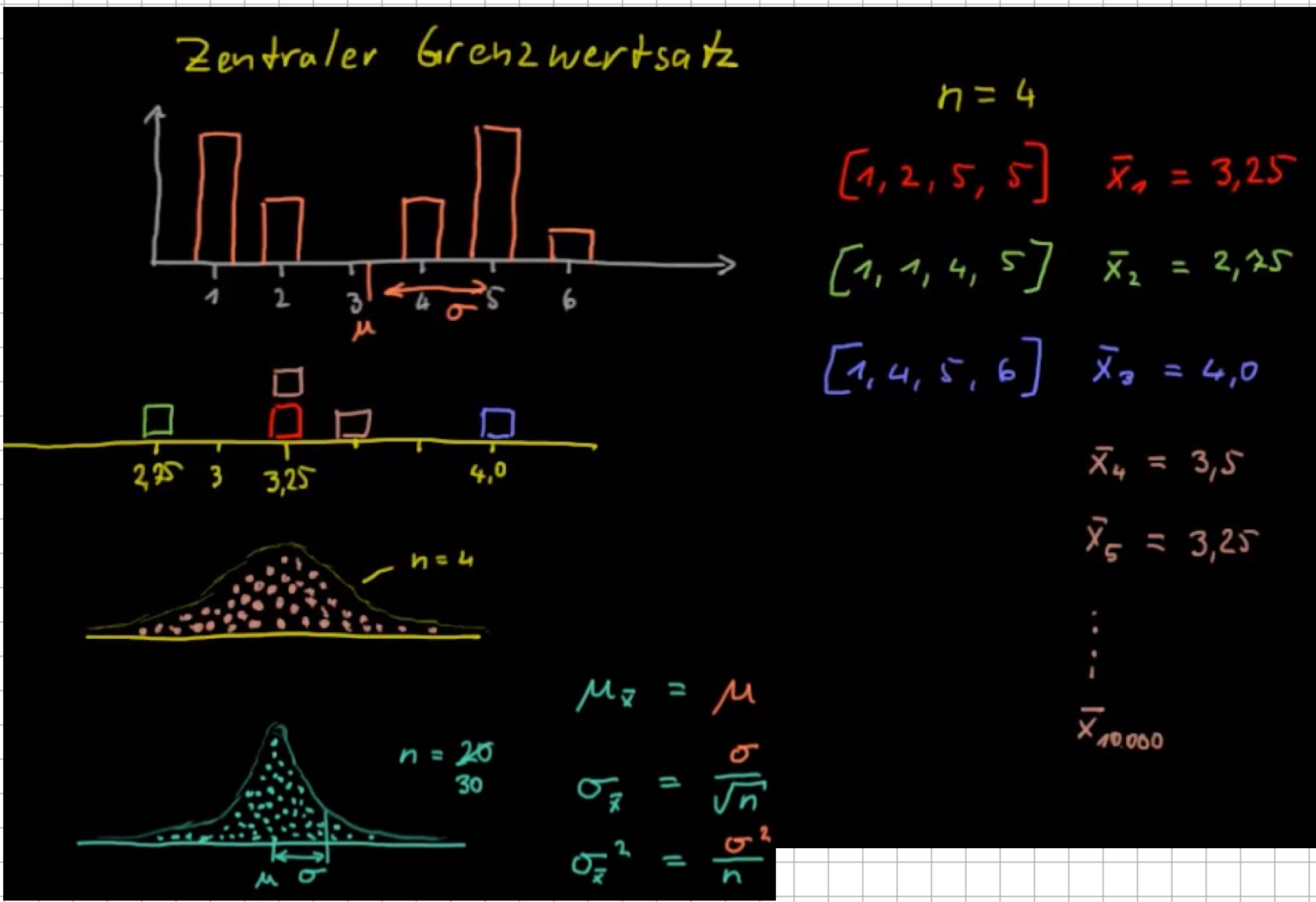
Gesuchter Wert	Grundgesamtheit	Stichprobe
Mittelwert	$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$	$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
Anteilswert	$\Theta = \frac{M}{N}$	$p = \frac{m}{n}$
Varianz	$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i-\mu_\alpha)^2$	$s^{*2} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$
Standardabweichung	$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu_\alpha)^2}$	$s^* = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}$

$$\tilde{s}^2 = \frac{\sum(x_i-\bar{x})^2}{n} = s^2 \cdot \frac{n-1}{n}$$

$$s^2 = \frac{\sum(x_i-\bar{x})^2}{n-1} = \tilde{s}^2 \cdot \frac{n}{n-1}$$

Zentraler Grenzwertsatz

Grundsatz des Zentralen Grenzwertsatzes:
Wir nehmen eine beliebige Verteilung, diskret oder stetig, und nehmen Stichproben. Für jede Stichprobe wird der Mittelwert ermittelt und tragen ihn in eine Häufigkeitsverteilung ein. Je grösser der Stichprobenumfang n gewählt wird, desto steiler wird die Normalverteilungskurve und desto genauer ist sie einer perfekten Normalvertielungskurve.



Mindestanzahl von Stichproben für eine Normalverteilung = 30

μ Mittelwert der Ausgangsverteilung (Grundgesamtheit) “Müh”

$\mu_{\bar{x}}$ Mittelwert der Stichproben-Mittelwerte —> Ist erwartungsgetreu zu “Müh”

σ Standardabweichung der Ausgangsverteilung “Sigma”

$\sigma_{\bar{x}}$ Standardabweichung der Stichproben-Mittelwerte

$\sigma_{\bar{x}}^2$ Varianz der Stichproben-Mittelwerte

$$\mu_{\bar{x}} = \mu$$
$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$
$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$$

Frage zu Grenzwertsatz:
Wie hoch ist die Wahrscheinlichkeit für ein en Wert inherhalb von 2 Standardabweichungen?

Es gibt praktisch unendlich viele Normalverteilungen, weil unendlich viele verschiedene Werte für μ und σ eingesetzt werden können. Allerdings kommt man trotzdem mit der Tabelle zur Standardnormalverteilung aus, denn alle Normalverteilungen lassen sich durch Standardisierung (F.4.6.8) auf die Standardnormalverteilung zurückführen:

$$F_N\left(\frac{x-\mu}{\sigma}\middle|\,\mu;\sigma\right) = F_N(z)$$

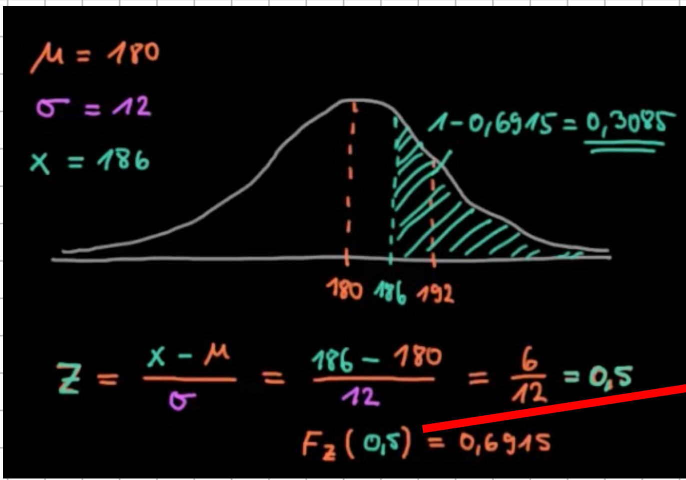
Was bedeutet diese Formel?:

- Nach dieser Formel bestimmt man Wahrscheinlichkeiten dafür, dass eine normalverteilte Zufallsvariable mit Zentrum bei μ (z.B. $\mu = 15$) und einer Standardabweichung von σ (z.B. $\sigma = 8$) höchstens den Wert x (also z.B. $x = 19$) annimmt.
- Man subtrahiert dazu μ vom x-Wert und dividiert das Ergebnis durch σ (also z.B. $(19-15)/8 = 0{,}5$).
- Das Ergebnis ist der z-Wert der Standardnormalverteilung, der dem x-Wert der gegebenen Normalverteilung entspricht. Deswegen kann man die gesuchte Wahrscheinlichkeit jetzt direkt aus der Tabelle zur Standardnormalverteilung ablesen (z.B. $F_N(z = 0{,}5) = 0{,}6915$).

Aufgabe 1 zum Zentralen Grenzwertsatz

Angenommen die Körpergröße erwachsener Männer in Deutschland sei normalverteilt mit einem Erwartungswert von 180 cm und einer Standardabweichung von 12 cm.

Bestimme die Wahrscheinlichkeit, dass ein zufällig ausgewählter Mann größer als 186 cm ist.

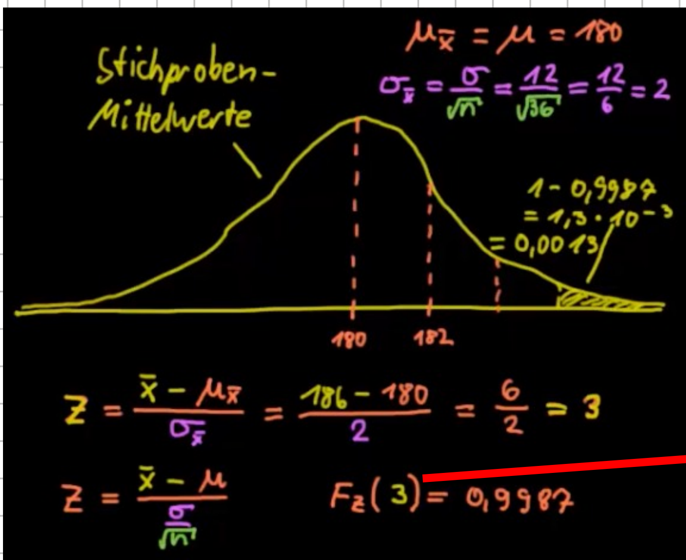


Die Wahrscheinlichkeit ist 30%

z	0,00
0,0	0,5000
0,1	0,5398
0,2	0,5793
0,3	0,6179
0,4	0,6554
0,5	0,6915

0.5 = ein halbe Standardabweichung entfernt

Bestimme die Wahrscheinlichkeit, dass der Mittelwert von 36 zufällig ausgewählten Männern mehr als 186 cm ist.



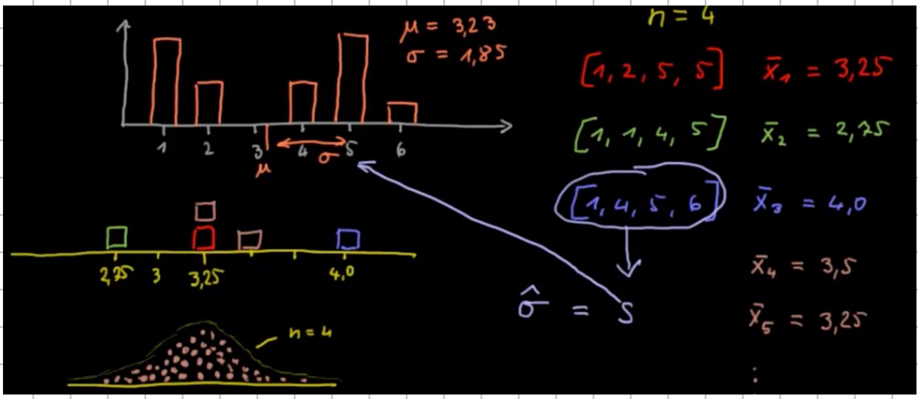
Die Wahrscheinlichkeit ist 0.13%

z	0,00
0,0	0,5000
0,1	0,5398
0,2	0,5793
2,8	0,9974
2,9	0,9981
3,0	0,9987

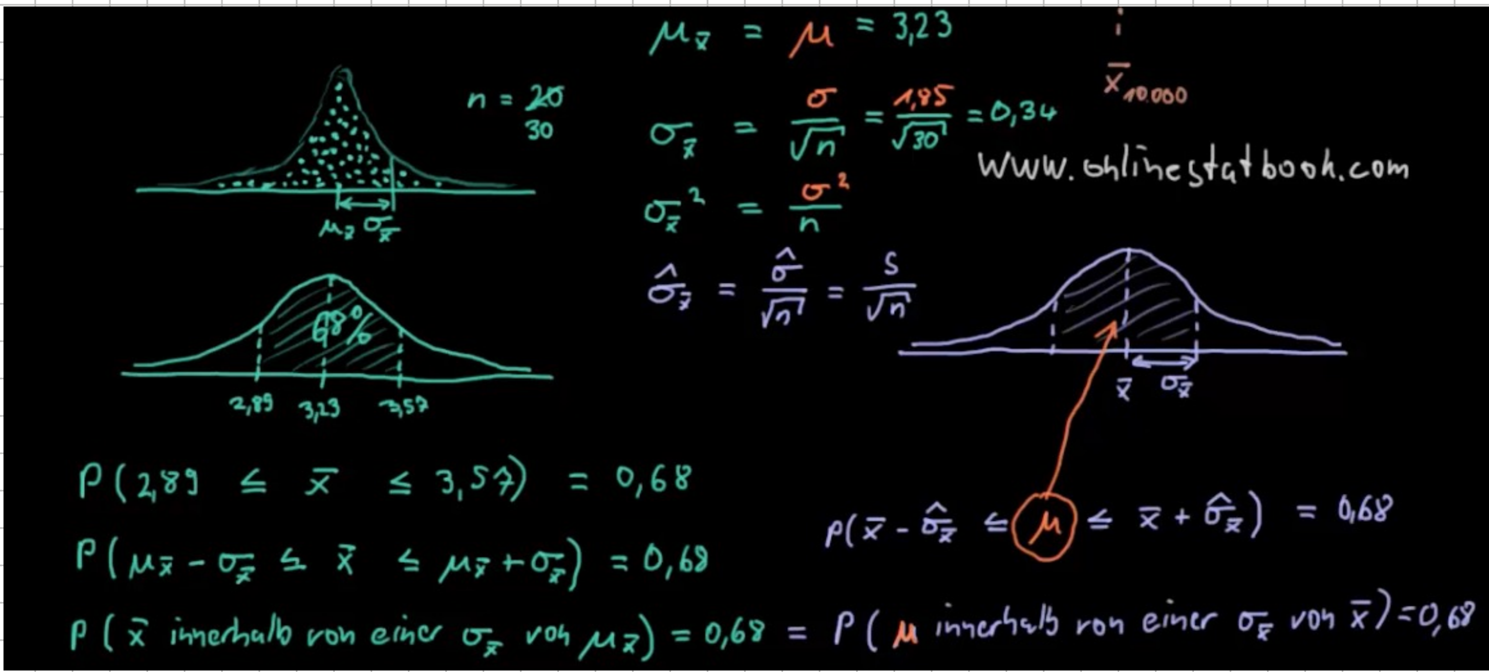
3 = drei Standardabweichungen entfernt

Zentraler Grenzwertsatz - Wahrscheinlichkeiten

Mindestanzahl von Stichproben für eine Normalverteilung = 30



Wenn die Standardabweichung der Ausgangsverteilung nicht bekannt ist kann sie Standardabweichung S einer Stichprobe verwendet werden.

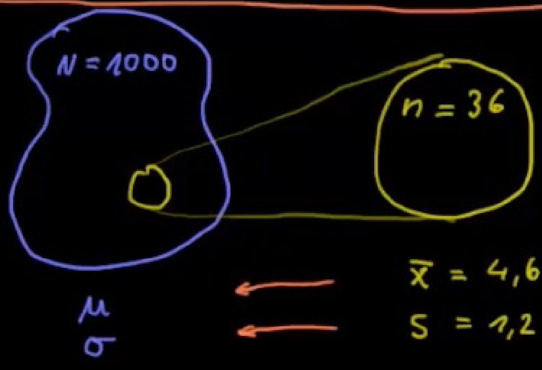


Aufgabe 2 zu Zentraler Grenzwertsatz

Ein LKW mit einer Ladekapazität von 5 Tonnen soll 1000 Melonen vom Hamburger Hafen zum Logistikzentrum eines Großhändlers in Hannover transportieren. Beim Wiegen von 36 Melonen ergab sich ein durchschnittliches Gewicht von 4,6 kg pro Melone bei einer Standardabweichung von S = 1,2 kg.
Wie groß ist die Wahrscheinlichkeit, dass der LKW mit den 1000 Melonen überladen ist? $\Rightarrow$ 2,5 %

$$\hat{s}^2 = \frac{\sum (x_i - \bar{x})^2}{n} = s^2 \cdot \frac{n-1}{n}$$

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \hat{s}^2 \cdot \frac{n}{n-1}$$



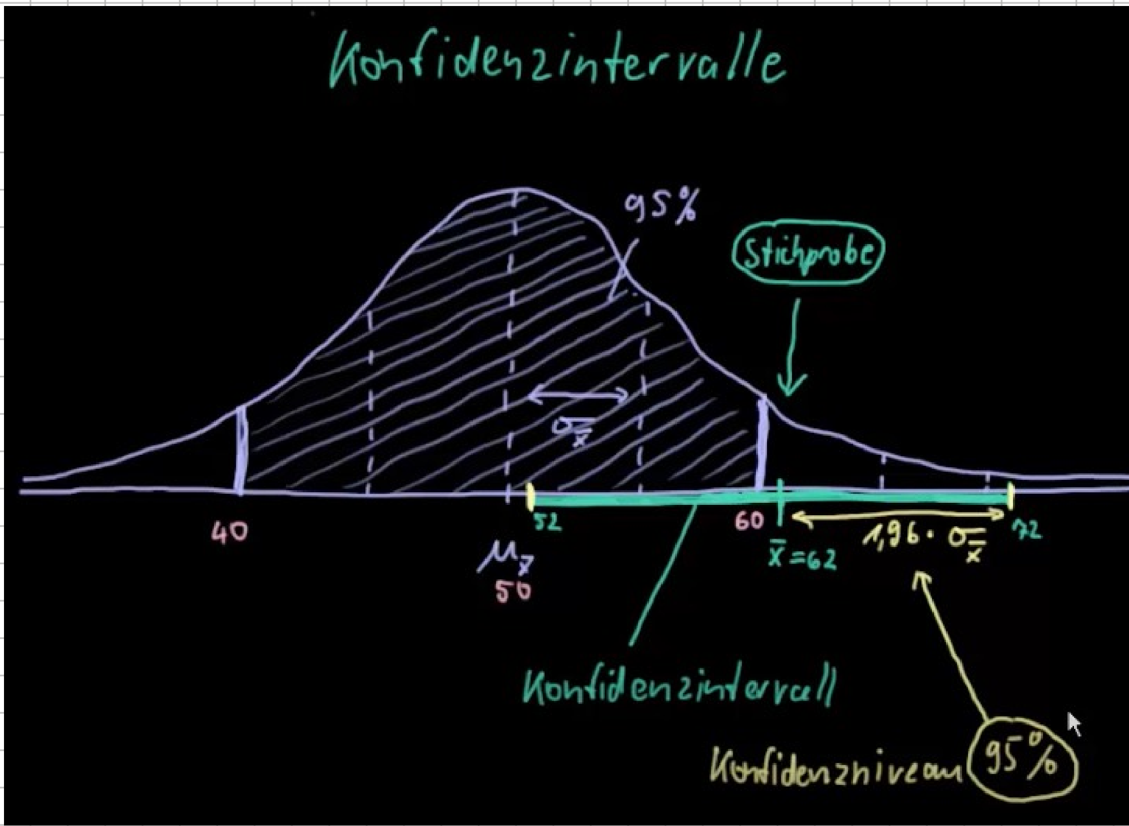
μ σ $\mu_{\bar{x}}$ $\sigma_{\bar{x}} = \frac{S}{\sqrt{n}} = \frac{1,2}{6} = 0,2$

Konfidenzintervalle

Ein Konfidenzintervall ist der Bereich für eine bestimmte Wahrscheinlichkeit

Konfidenz —> Zuversicht

Frage zu Konfidenzintervallen:
Wie ist der Bereich der Standardabweichung wenn die Wahrscheinlichkeit xx% ist



Mittelwert der Stichprobe

Es wird ein Konfidenziintervall zu einem zugehörigen Konzidenzniveau gesucht.

Konfidenzintervall

Konfidenzniveau

Irrtumswahrscheinlichkeit

Konfidenzwahrscheinlichkeit

Die Prüfung von 50 Shampoo-Flaschen ergab eine mittlere Füllmenge von 308ml mit einer Standardabweichung S von 10ml.

Bestimme das zweiseitige Konfidenzintervall [μ_u ; μ_o] zum Konfidenzniveau $\boxed{1-\alpha = 0{,}9}$

Bestimme das einseitige Konfidenzintervall [$-\infty$; μ_o] zum Konfidenzniveau $\boxed{1-\alpha = 0{,}9}$

$F_Z(z) = 0{,}9$

$F_Z(z) = 0{,}9$

$KI = [U ; O]$

$U = \bar{x} - z_{0,95} \cdot \sigma_{\bar{x}} = \bar{x} - 1{,}65\,\sigma_{\bar{x}} = \bar{x} - z_{1-\frac{\alpha}{2}} \cdot \sigma_{\bar{x}}$

$O = \bar{x} + z_{0,95} \cdot \sigma_{\bar{x}} = \bar{x} + 1{,}65\,\sigma_{\bar{x}} = \bar{x} + z_{1-\frac{\alpha}{2}} \cdot \sigma_{\bar{x}}$

$KI = [-\infty ; O]$

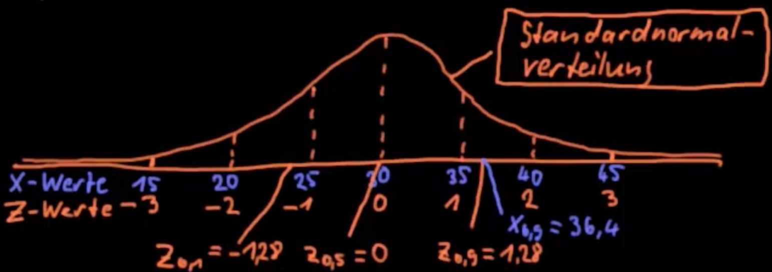
$O = \bar{x} + z_{0,9} \cdot \sigma_{\bar{x}} = \bar{x} + 1{,}28\,\sigma_{\bar{x}} = \bar{x} + z_{1-\alpha} \cdot \sigma_{\bar{x}}$

$F_Z(1) =$  $= 0{,}68$

$F_Z(1) =$  $= 0{,}84$

21, 23, 23, 23, 24, 25, 27, 27, 32, 45

$x_{0,1} = 22$ $x_{0,4} = 23{,}5$ $x_{0,75} = 27$



$F_Z(1{,}28) = 0{,}9$

$z_{0,9} = 1{,}28$

$F_Z(z_{0,9}) = 0{,}9$

$F_Z(z_{0,5}) = 0{,}5$

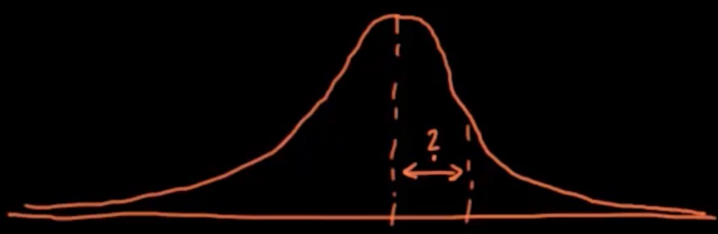
$z_{0,5} = 0$

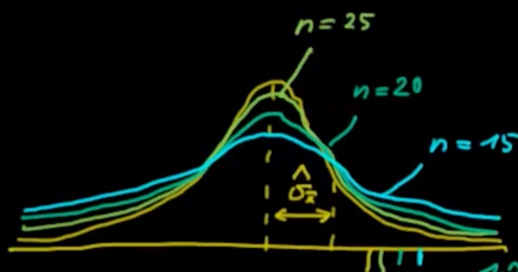
$F_Z(0) = 0{,}5$

Ein Shampoo-Hersteller verspricht, dass die Füllmenge seiner Flaschen normalverteilt sei.

Die Prüfung von 25 Shampoo-Flaschen ergab eine mittlere Füllmenge von 308ml mit einer Standardabweichung S von 12ml.

Bestimme das zweiseitige Konfidenzintervall [μ_u ; μ_o] zum Konfidenzniveau $1-\alpha = 0,9$





$z_{0,95}$ $t_{0,95} = \cancel{1,708}$ 1,711

$t_{0,95}(25) = 1,708$

$t_{0,95}(24) = 1,711$

$t_{0,95}(n-1) = 1,711$

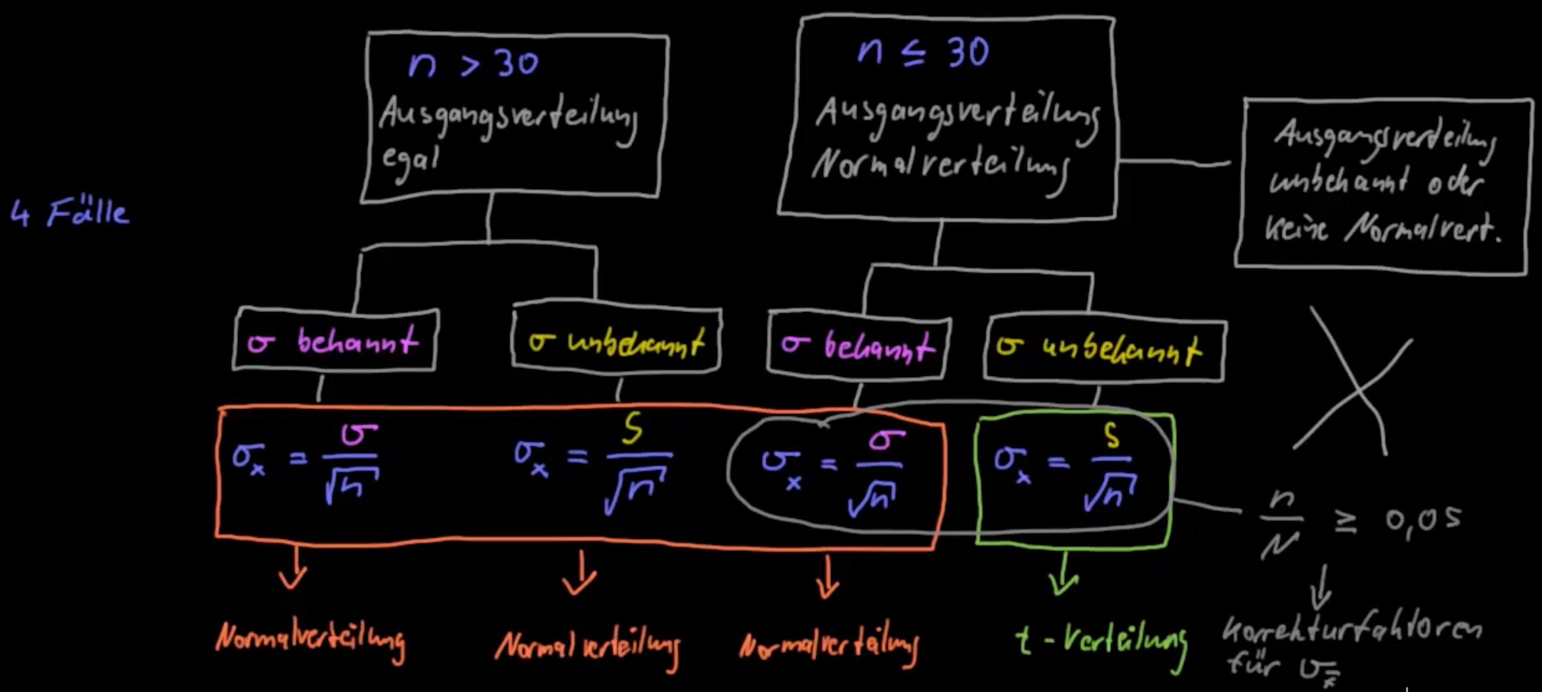
t - Verteilung

$$\sigma_{\bar{x}} = \frac{\hat{\sigma}}{\sqrt{n}} = \frac{s}{\sqrt{n}} = \frac{12}{5} = 2,4$$

$$U = \bar{x} - t_{0,95} \cdot \sigma_{\bar{x}} = 308 - \cancel{1,708}\,1,711 \cdot 2,4 = 303,9$$

$$O = \bar{x} + t_{0,95} \cdot \sigma_{\bar{x}} = 308 + \cancel{1,708}\,1,711 \cdot 2,4 = 312,1$$

Konfidenzintervalle zusammengefasst



Z-Werte der Standardnormalverteilung

p	0,500	0,600	0,700	0,800	0,900	0,950	0,975	0,990	0,995	0,999
z_p	0,0000	0,2533	0,5244	0,8416	1,2816	1,6449	1,9600	2,3263	2,5758	3,0902

Quantile z_p der Standardnormalverteilung

z	0	1	2	3	4	5	6	7	8	9
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8079	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9956	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998
3,5	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998
3,6	0,9998	0,9998	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,7	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,8	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

Werte der Verteilungsfunktion $\Phi(z)$ der Standardnormalverteilung

T-Werte der T-Verteilung

Für den n-Wert wird in der T-Werte-Tabelle immer n-1 abgelesen!

n-1

ν	0,800	0,850	0,900	0,950	0,975	0,990	0,995
1	1,376	1,963	3,078	6,314	12,706	31,821	63,657
2	1,061	1,386	1,886	2,920	4,303	6,965	9,925
3	0,979	1,250	1,638	2,353	3,182	4,541	5,841
4	0,941	1,190	1,533	2,132	2,776	3,747	4,604
5	0,920	1,156	1,476	2,015	2,571	3,365	4,032
6	0,906	1,134	1,440	1,943	2,447	3,143	3,707
7	0,896	1,119	1,415	1,895	2,365	2,998	3,499
8	0,889	1,108	1,397	1,860	2,306	2,896	3,355
9	0,883	1,100	1,383	1,833	2,262	2,821	3,250
10	0,879	1,093	1,372	1,812	2,228	2,764	3,169
11	0,876	1,088	1,363	1,796	2,201	2,718	3,106
12	0,873	1,083	1,356	1,782	2,179	2,681	3,055
13	0,870	1,080	1,350	1,771	2,160	2,650	3,012
14	0,868	1,076	1,345	1,761	2,145	2,624	2,977
15	0,866	1,074	1,341	1,753	2,131	2,602	2,947
16	0,865	1,071	1,337	1,746	2,120	2,583	2,921
17	0,863	1,069	1,333	1,740	2,110	2,567	2,898
18	0,862	1,067	1,330	1,734	2,101	2,552	2,878
19	0,861	1,066	1,328	1,729	2,093	2,539	2,861
20	0,860	1,064	1,325	1,725	2,086	2,528	2,845
21	0,859	1,063	1,323	1,721	2,080	2,518	2,831
22	0,858	1,061	1,321	1,717	2,074	2,508	2,819
23	0,858	1,060	1,319	1,714	2,069	2,500	2,807
24	0,857	1,059	1,318	1,711	2,064	2,492	2,797
25	0,856	1,058	1,316	1,708	2,060	2,485	2,787
26	0,856	1,058	1,315	1,706	2,056	2,479	2,779
27	0,855	1,057	1,314	1,703	2,052	2,473	2,771
28	0,855	1,056	1,313	1,701	2,048	2,467	2,763
29	0,854	1,055	1,311	1,699	2,045	2,462	2,756
30	0,854	1,055	1,310	1,697	2,042	2,457	2,750
31	0,853	1,054	1,310	1,696	2,040	2,4528	2,744
32	0,853	1,054	1,309	1,694	2,074	2,4587	2,739
33	0,853	1,053	1,308	1,692	2,069	2,4448	2,733
34	0,852	1,053	1,307	1,691	2,064	2,4411	2,728
35	0,852	1,052	1,306	1,690	2,060	2,4477	2,724
36	0,852	1,052	1,306	1,688	2,056	2,4345	2,720
37	0,851	1,051	1,305	1,687	2,052	2,4314	2,715
38	0,851	1,051	1,304	1,686	2,048	2,4386	2,712
39	0,851	1,050	1,304	1,685	2,045	2,4258	2,708
40	0,851	1,050	1,303	1,684	2,021	2,4233	2,705

Quantile der t-Verteilung

Kontingenztabellen

STATISTISCHE EINHEIT	SCHIRM
MERKMAL 1	JAHRESZEIT
MERKMAL 2	FARBE

ZWEIDIMENSIONALER DATENSATZ

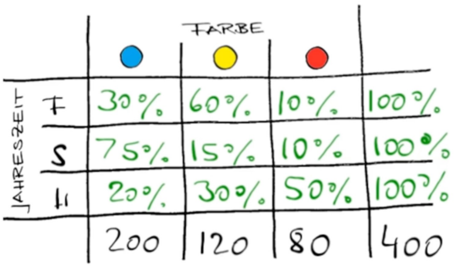
GEMEINSAME EMPIRISCHE HÄUFIGKEITEN

RANDHÄUFIGKEITEN

KONTINGENZTABELLE

0.075	0.15	0.125	100
0.375	0.25	0.25	200
0.05	0.25	0.625	100
100%	100%	100%	400

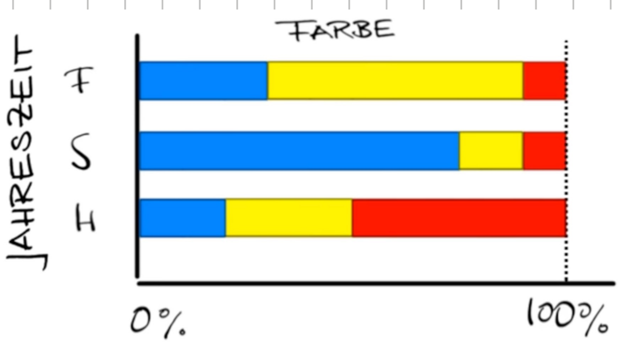
Relative Häufigkeit auf Jahreszeit bezogen



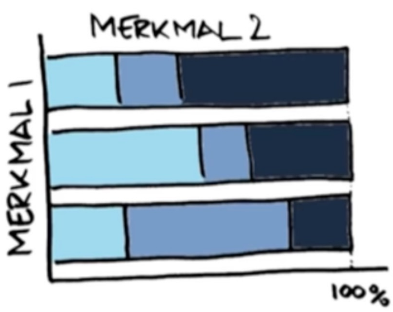
Relative Häufigkeit auf Farben bezogen

0.15	0.5	0.125	100
0.75	0.25	0.25	200
0.10	0.25	0.625	100
100%	100%	100%	400

Streifendiagramm



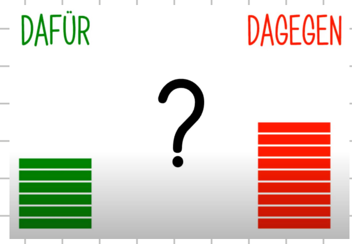
STREIFENDIAGRAMM

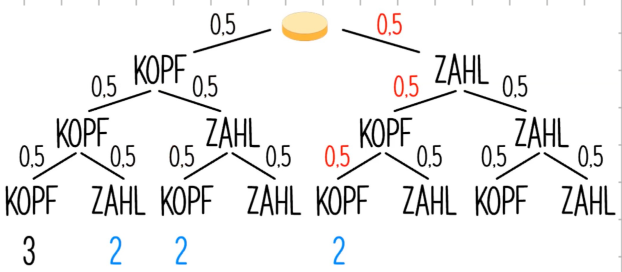


STREIFEN-DIAGRAMM

Binomialtest

Der Binomialtest kann eingesetzt werden, wenn man Hypothesen zu Anteilswerten prüfen möchte (z.B.: „Weniger als 50% der Kinogänger mögen salziges Popcorn.").





$0{,}5^3 + 0{,}5^3 + 0{,}5^3$

$$\binom{n}{k} \cdot \Theta^k \cdot (1-\Theta)^{n-k}$$

n: ANZAHL SCHRITTE

Θ: WAHRSCHEINLICHKEIT JE SCHRITT

k: ANZAHL "ERFOLGE"

BINOMIALVERTEILUNG

$$\binom{60}{k} \cdot 0{,}5^k \cdot (1-0{,}5)^{60-k}$$

n: 60 KUNDEN

Θ: 0,5 (MATERIAL EGAL)

0: ANZAHL "PAPIERTÜTENBEVORZUGER"

$$\binom{60}{0} \cdot 0{,}5^0 \cdot (1-0{,}5)^{60-0} \approx 8{,}7 \cdot 10^{-19}$$

8.7×10^{-19}

+

$$\binom{60}{k} \cdot 0{,}5^k \cdot (1-0{,}5)^{60-k}$$

n: 60 KUNDEN

Θ: 0,5 (MATERIAL EGAL)

1: ANZAHL "PAPIERTÜTENBEVORZUGER"

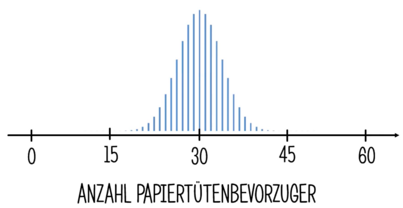
$$\binom{60}{1} \cdot 0{,}5^1 \cdot (1-0{,}5)^{60-1} \approx 5{,}2 \cdot 10^{-17}$$

5.2×10^{-17}

+

⋮

HYPOTHESE: MATERIAL EGAL $\Rightarrow \Theta = 0{,}5$

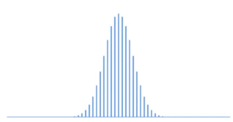


$1-\alpha$ = 1- 0.05 = 0.95

Beim x-ten Versuch der die Summe der Wahrscheinlichkeiten über die 0.95 bringt stopt der Test. Dieser Wert ist der p-Wert.

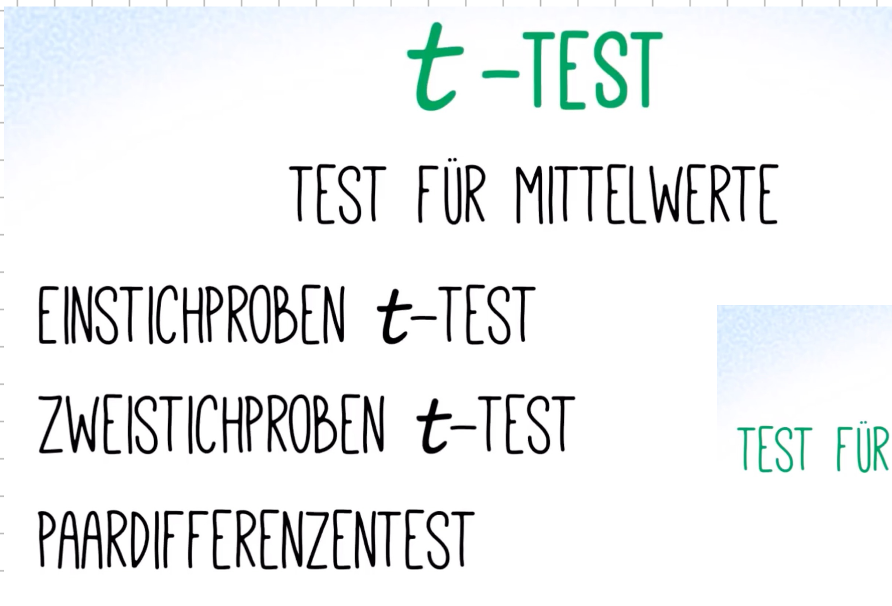
100% - p-Wert in % = % Ablehnungsbereich

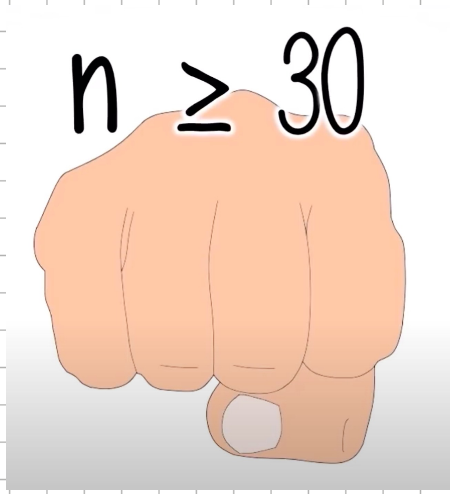
SACHVERHALT	"MATERIAL EGAL"
HYPOTHESE $\mathcal{H}_0$	$\Theta = 0{,}5$
TESTVERTEILUNG	BINOMIALVERTEILUNG

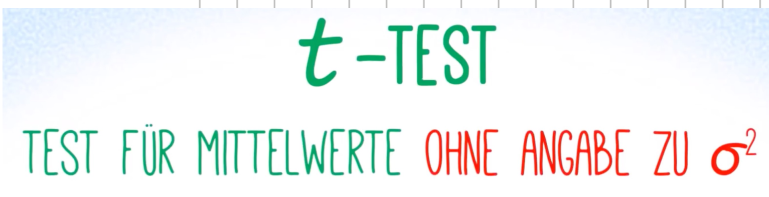


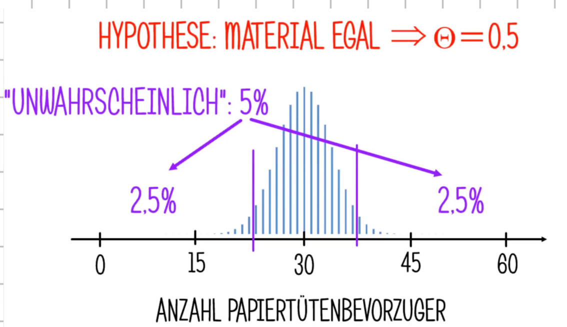
Statistische Tests

t-Test

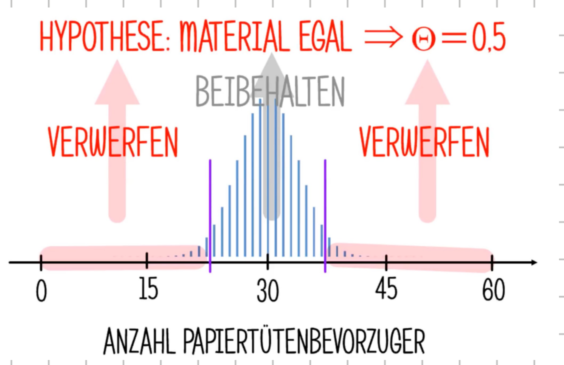


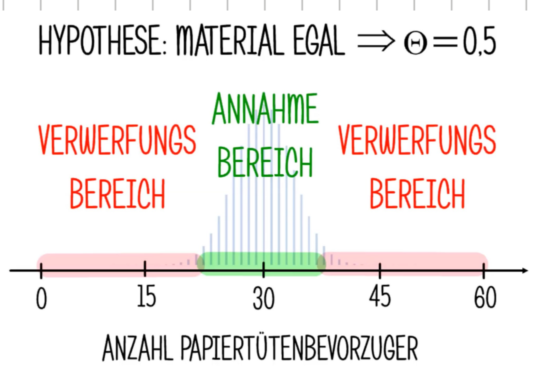


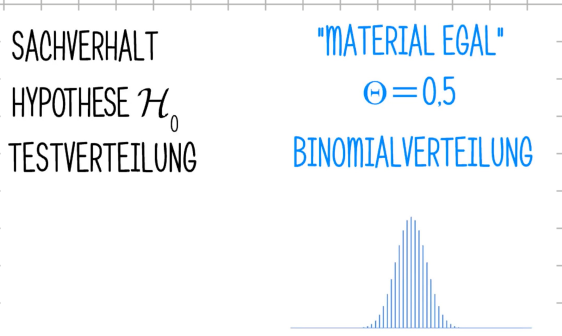




Sowohl der t-Test als auch die Varianzanalyse dienen dazu, Hypothesen zu Mittelwerten zu prüfen.
t-Test
Den t-Test verwendet man, wenn es dabei um eine Gesamtheit (z.B. „Die durchschnittliche Zahlungsbereitschaft für Popcorn liegt unter 3,- EUR/100g.") oder zwei Gesamtheiten geht (z.B. „Die durchschnittliche Zahlungsbereitschaft für Popcorn ist in der Stadt um 0,50 EUR/100g höher als auf dem Land.").
Varianzanalyse
Bei drei oder mehr Gesamtheiten verwendet man die Varianzanalyse (z.B. „Die Zahlungsbereitschaft für Popcorn unterscheidet sich bei verschiedenen Filmgenres (Komödie, Liebesfilm, Thriller usw.) deutlich voneinander.").







EINSTICHPROBEN t-TEST

SACHVERHALT: ANGABE IM DURCHSCHNITT KORREKT?

HYPOTHESE H_0: $\mu = 10G/100ML$

TESTVERTEILUNG: t-VERTEILUNG

TESTNIVEAU: $\alpha = 0{,}05$

TESTSTATISTIK: $t = \dfrac{\overline{x} - \mu}{\sqrt{s^{*2}/n}}$ $\quad t = \dfrac{10{,}5 - 10}{\sqrt{0{,}64/10}} \approx 1{,}98$

Stichprobenvarianz

$$s^{*2} = \frac{\sum (x_i - \overline{x})^2}{n - 1}$$

$$\frac{10 + 9{,}2 + 10{,}9 + 11{,}3 + 10{,}5 + 10{,}9 + 10{,}6 + 11{,}5 + 9{,}2 + 10{,}9}{10}$$

$\overline{x} = 10{,}5$ G/100ML

10 FLASCHEN 1 STREUPARAMETER

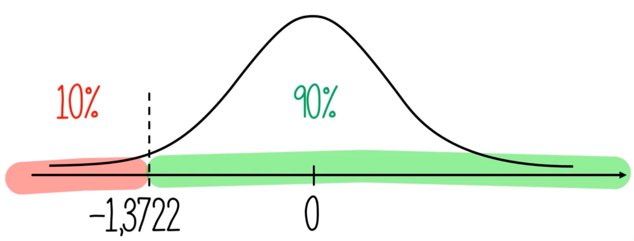
10 − 1 = 9 FREIHEITSGRADE

10 9,2 10,9 11,3 10,5 10,9 10,6 11,5 9,2 10,9

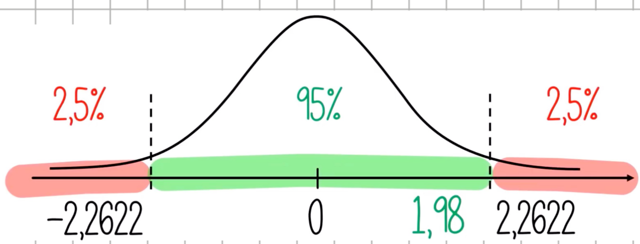
$$s^{*2} = \frac{(10 - 10{,}5)^2 + (9{,}2 - 10{,}5)^2 + \dots + (10{,}9 - 10{,}5)^2}{9}$$

$$= 0{,}64$$

Einseitig bei $\alpha = 10\%$



Zweiseitig bei $\alpha = 5\%$



Anhang C: Tabelle zur t-Verteilung

Werte gelten für den Bereich Rechts! Für Links x(-1)

	einseitig				zweiseitig (symmetrische Intervalle)			
α	0,1	0,05	0,01	0,001	0,1	0,05	0,01	0,001
$1-\alpha$	0,9	0,95	0,99	0,999	0,9	0,95	0,99	0,999
1	3,0777	6,3138	31,821	318,31	6,3138	12,706	63,657	636,62
2	1,8856	2,9200	6,9646	22,327	2,9200	4,3027	9,9248	31,599
3	1,6377	2,3534	4,5407	10,215	2,3534	3,1824	5,8409	12,924
4	1,5332	2,1318	3,7469	7,1732	2,1318	2,7764	4,6041	8,6103
5	1,4759	2,0150	3,3649	5,8934	2,0150	2,5706	4,0321	6,8688
6	1,4398	1,9432	3,1427	5,2076	1,9432	2,4469	3,7074	5,9588
7	1,4149	1,8946	2,9980	4,7853	1,8946	2,3646	3,4995	5,4079
8	1,3968	1,8595	2,8965	4,5008	1,8595	[illegible]	3,3554	5,0413
9	1,3830	1,8331	2,8214	4,2968	1,8331	$\boxed{2{,}2622}$	3,2498	4,7809
10	1,3722	1,8125	2,7638	4,1437	1,8125	2,2281	3,1693	4,5869

ZWEISTICHPROBEN t-TEST

SACHVERHALT: MINDESTENS 800,– HÖHER?

HYPOTHESE H_0: $\mu_{DESIGN} - \mu_{NORMAL} \geq 800,-$

TESTVERTEILUNG: t-VERTEILUNG, $n_1 + n_2 - 2$ FG

TESTNIVEAU: $\alpha = 0{,}1$

TESTSTATISTIK: $t \approx -1{,}73$

$$\begin{array}{r} 2.900 \\ +\ 2.400 \\ +\ 2.700 \\ +\ 2.400 \\ +\ 2.700 \\ +\ 2.500 \\ \hline 15.600 \end{array} \qquad 15.600 : 6 = 2.600$$

$$12.000 : 6 = 2.000 \qquad \begin{array}{r} 2.100 \\ +\ 1.800 \\ +\ 1.900 \\ +\ 1.800 \\ +\ 2.100 \\ +\ 2.300 \\ \hline 12.000 \end{array}$$

SACHVERHALT: MINDESTENS 800,– HÖHER?

HYPOTHESE H_0: $\mu_{DESIGN} - \mu_{NORMAL} \geq 800,-$

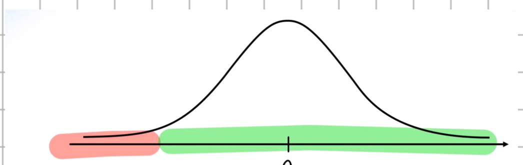
$$\begin{array}{r} +\ (2.900 - 2.600)^2 \\ +\ (2.400 - 2.600)^2 \\ +\ (2.700 - 2.600)^2 \\ +\ (2.400 - 2.600)^2 \\ +\ (2.700 - 2.600)^2 \\ +\ (2.500 - 2.600)^2 \\ \hline 200.000 \end{array} \qquad 200.000 : 5 = 40.000$$

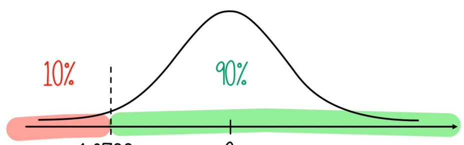
$$200.000 : 5 = 40.000 \qquad \begin{array}{r} +\ (2.100 - 2.000)^2 \\ +\ (1.800 - 2.000)^2 \\ +\ (1.900 - 2.000)^2 \\ +\ (1.800 - 2.000)^2 \\ +\ (2.100 - 2.000)^2 \\ +\ (2.300 - 2.000)^2 \\ \hline 200.000 \end{array}$$

$$t = \frac{\overline{x}^{(1)} - \overline{x}^{(2)} - \omega_0}{\sqrt{\dfrac{(n_1-1)\cdot s^{(1)*2} + (n_2-1)\cdot s^{(2)*2}}{n_1+n_2-2}}} \cdot \sqrt{\frac{n_1 \cdot n_2}{n_1+n_2}}$$

$$t = \frac{2.600 - 2.000 - 800}{\sqrt{\dfrac{(6-1)\cdot 40.000 + (6-1)\cdot 40.000}{6+6-2}}} \cdot \sqrt{\frac{6\cdot 6}{6+6}} \approx -1{,}73$$

HYPOTHESE H_0: $\mu_{DESIGN} - \mu_{NORMAL} \geq 800,-$





Anhang C: Tabe

	Werte
α / $1-\alpha$	0,1 / 0,9
1	3,0777
2	1,8856
3	1,6377
4	1,5332
5	1,4759
6	1,4398
7	1,4149
8	1,3968
9	1,3830
10	$\boxed{1{,}3722}$

PAARDIFFERENZENTEST

SACHVERHALT: ERHÖHT UV DIE HALTBARKEIT?

HYPOTHESE H_0: $\mu_d \leqq 0$ TAGE

TESTVERTEILUNG: t-VERTEILUNG, 5 FREIHEITSGRADE

TESTNIVEAU: $\alpha = 0{,}05$

TESTSTATISTIK:

$$t = \frac{\bar{d} - \omega_0}{\sqrt{s^{*2}/n}} \qquad t = \frac{1{,}5 - 0}{\sqrt{5{,}5 / 6}} \approx 1{,}57$$

	MIT		OHNE	DIFFERENZ
	5	-	6	− 1
	9	-	8	+ 1
	8	-	9	− 1
	17	-	14	+ 3
	22	-	20	+ 2
	45	-	40	+ 5
				9

$$9 : 6 = 1{,}5$$

	MIT		OHNE	DIFFERENZ
	5	-	6	$(-1-1{,}5)^2$
	9	-	8	$(+1-1{,}5)^2$
	8	-	9	$(-1-1{,}5)^2$
	17	-	14	$(+3-1{,}5)^2$
	22	-	20	$(+2-1{,}5)^2$
	45	-	40	$(+5-1{,}5)^2$
				27,5

$$27{,}5 : 5 = 5{,}5$$

Anhang C: Tabelle zur *t*-Verteilung

Werte gelten für den Bereich Rechts!
Für Links x(-1)

	einseitig				zweiseitig (symmetrische Intervalle)			
α	0,1	0,05	0,01	0,001	0,1	0,05	0,01	0,001
$1-\alpha$	0,9	0,95	0,99	0,999	0,9	0,95	0,99	0,999
Freiheitsgrade								
1	3,0777	6,3138	31,821	318,31	6,3138	12,706	63,657	636,62
2	1,8856	2,9200	6,9646	22,327	2,9200	4,3027	9,9248	31,599
3	1,6377	2,3534	4,5407	10,215	2,3534	3,1824	5,8409	12,924
4	1,5332	2,1318	3,7469	7,1732	2,1318	2,7764	4,6041	8,6103
5	1,4759	2,0150	3,3649	5,8934	2,0150	2,5706	4,0321	6,8688
6	1,4398	1,9432	3,1427	5,2076	1,9432	2,4469	3,7074	5,9588
7	1,4149	1,8946	2,9980	4,7853	1,8946	2,3646	3,4995	5,4079
8	1,3968	1,8595	2,8965	4,5008	1,8595	2,3060	3,3554	5,0413
9	1,3830	1,8331	2,8214	4,2968	1,8331	2,2622	3,2498	4,7809
10	1,3722	1,8125	2,7638	4,1437	1,8125	2,2281	3,1693	4,5869
11	1,3634	1,7959	2,7181	4,0247	1,7959	2,2010	3,1058	4,4370
12	1,3562	1,7823	2,6810	3,9296	1,7823	2,1788	3,0545	4,3178
13	1,3502	1,7709	2,6503	3,8520	1,7709	2,1604	3,0123	4,2208
14	1,3450	1,7613	2,6245	3,7874	1,7613	2,1448	2,9768	4,1405
15	1,3406	1,7531	2,6025	3,7328	1,7531	2,1314	2,9467	4,0728
16	1,3368	1,7459	2,5835	3,6862	1,7459	2,1199	2,9208	4,0150
17	1,3334	1,7396	2,5669	3,6458	1,7396	2,1098	2,8982	3,9651
18	1,3304	1,7341	2,5524	3,6105	1,7341	2,1009	2,8784	3,9216
19	1,3277	1,7291	2,5395	3,5794	1,7291	2,0930	2,8609	3,8834
20	1,3253	1,7247	2,5280	3,5518	1,7247	2,0860	2,8453	3,8495
21	1,3232	1,7207	2,5176	3,5272	1,7207	2,0796	2,8314	3,8193
22	1,3212	1,7171	2,5083	3,5050	1,7171	2,0739	2,8188	3,7921
23	1,3195	1,7139	2,4999	3,4850	1,7139	2,0687	2,8073	3,7676
24	1,3178	1,7109	2,4922	3,4668	1,7109	2,0639	2,7969	3,7454
25	1,3163	1,7081	2,4851	3,4502	1,7081	2,0595	2,7874	3,7251
26	1,3150	1,7056	2,4786	3,4350	1,7056	2,0555	2,7787	3,7066
27	1,3137	1,7033	2,4727	3,4210	1,7033	2,0518	2,7707	3,6896
28	1,3125	1,7011	2,4671	3,4082	1,7011	2,0484	2,7633	3,6739
29	1,3114	1,6991	2,4620	3,3962	1,6991	2,0452	2,7564	3,6594
30	1,3104	1,6973	2,4573	3,3852	1,6973	2,0423	2,7500	3,6460
31	1,3095	1,6955	2,4528	3,3749	1,6955	2,0395	2,7440	3,6335
32	1,3086	1,6939	2,4487	3,3653	1,6939	2,0369	2,7385	3,6218
33	1,3077	1,6924	2,4448	3,3563	1,6924	2,0345	2,7333	3,6109
34	1,3070	1,6909	2,4411	3,3479	1,6909	2,0322	2,7284	3,6007
35	1,3062	1,6896	2,4377	3,3400	1,6896	2,0301	2,7238	3,5911
36	1,3055	1,6883	2,4345	3,3326	1,6883	2,0281	2,7195	3,5821
37	1,3049	1,6871	2,4314	3,3256	1,6871	2,0262	2,7154	3,5737
38	1,3042	1,6860	2,4286	3,3190	1,6860	2,0244	2,7116	3,5657
39	1,3036	1,6849	2,4258	3,3128	1,6849	2,0227	2,7079	3,5581
40	1,3031	1,6839	2,4233	3,3069	1,6839	2,0211	2,7045	3,5510
100	1,2901	1,6602	2,3642	3,1737	1,6602	1,9840	2,6259	3,3905

Tabelle 79: Tabelle zur t-Verteilung

χ^2-Test

$F(x) = F_P(x|\lambda)$ $F(x) = F_N(x|\mu;\sigma)$ $F(\text{FARBEN}) = F_{\text{GLEICH}}(x)$

χ^2–ANPASSUNGSTEST χ^2–HOMOGENITÄTSTEST χ^2–UNABHÄNGIGKEITSTEST

χ^2–ANPASSUNGSTEST

SACHVERHALT: FARBEN GLEICHHÄUFIG VERKAUFT?

HYPOTHESE H_0: $F(\text{FARBEN}) = F_{\text{GLEICH}}(x)$

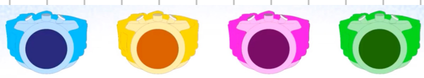
TESTVERTEILUNG: χ^2–VERTEILUNG

TESTNIVEAU: $\alpha = 5\%$

TESTSTATISTIK: $\chi^2_{EMP} = \sum \frac{(n_o - n_e)^2}{n_e}$

HYPOTHESE H_0: FARBEN WERDEN GLEICHHÄUFIG VERKAUFT

n_o	245	252	254	249	1.000
	OBSERVED				
n_e	250	250	250	250	1.000
	EXPECTED				



n_o	300	200	400	100	1.000
n_e	250	250	250	250	

$$\frac{(300-250)^2}{250} + \frac{(200-250)^2}{250} + \frac{(400-250)^2}{250} + \frac{(100-250)^2}{250}$$

TESTSTATISTIK $= 200 = \chi^2_{EMP}$

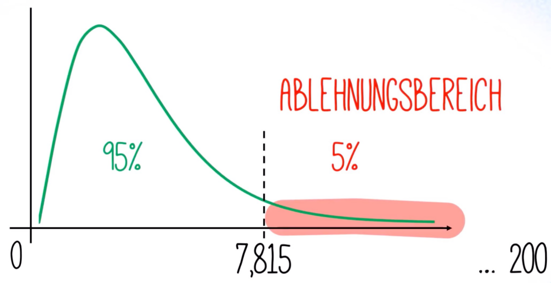
3 FREIHEITSGRADE

n_o	300	200	400	100	1.000
n_e	250	250	250	250	

Anhang B: Tabelle zur χ^2-Verteilung

α / $1-\alpha$	0,9 / 0,1	0,75 / 0,25	0,5 / 0,5	0,25 / 0,75	0,1 / 0,9	0,05 / 0,95
1	0,016	0,102	0,455	1,323	2,706	3,841
2	0,211	0,575	1,386	2,773	4,605	5,991
3	0,584	1,213	2,366	4,108	6,251	$\boxed{7,815}$

χ^2–VERTEILUNG



χ^2-HOMOGENITÄTSTEST

SACHVERHALT: FARBEN IM GLEICHEN VERHÄLTNIS?

HYPOTHESE H_0: $F_{ONLINE}(x) = F_{BRAND}(x) = F_{SPORT}(x)$

TESTVERTEILUNG: χ^2-VERTEILUNG

TESTNIVEAU: $\alpha = 10\%$

TESTSTATISTIK: $\chi^2_{EMP} = \sum \frac{(n_0 - n_e)^2}{n_e} = 8{,}75 = \chi^2_{EMP}$

n_0 n_e					
@	135 3/10 150	110 2/10 100	210 4/10 200	45 1/10 50	500
	105 3/10 90	54 2/10 60	108 4/10 120	33 1/10 30	300
	60 3/10 60	36 2/10 40	82 4/10 80	22 1/10 20	200
	300	200	400	100	1.000

n_0 n_e					
@	$\frac{(135-150)^2}{150}$	$+\frac{(110-100)^2}{100}$	$+\frac{(210-200)^2}{200}$	$+\frac{(45-50)^2}{50}$	500
	$+\frac{(105-90)^2}{90}$	$+\frac{(54-60)^2}{60}$	$+\frac{(108-120)^2}{120}$	$+\frac{(33-30)^2}{30}$	300
	$+\frac{(60-60)^2}{60}$	$+\frac{(36-40)^2}{40}$	$+\frac{(82-80)^2}{80}$	$+\frac{(22-20)^2}{20}$	200
	300	200	400	100	1.000

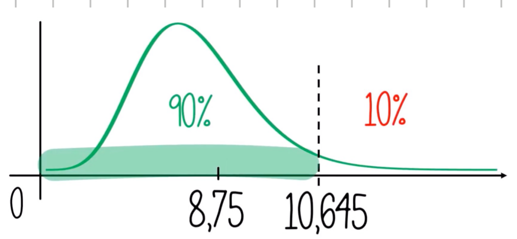
$= 8{,}75 = \chi^2_{EMP}$

6 FREIHEITSGRADE

135	110	210	45	500
105	54	108	33	300
60	36	82	22	200
300	200	400	100	1.000

Anhang B: Tabelle zur χ^2-Verteilung

α / $1-\alpha$	0,9 / 0,1	0,75 / 0,25	0,5 / 0,5	0,25 / 0,75	0,1 / 0,9
1	0,016	0,102	0,455	1,323	2,706
2	0,211	0,575	1,386	2,773	4,605
3	0,584	1,213	2,366	4,108	6,251
4	1,064	1,923	3,357	5,385	7,779
5	1,610	2,675	4,351	6,626	9,236
6	2,204	3,455	5,348	7,841	$\boxed{10{,}645}$



χ^2-UNABHÄNGIGKEITSTEST

SACHVERHALT: ABSATZSTEIGERUNG DURCH KAMPAGNE?

HYPOTHESE H_0: YOUTUBE-KAMPAGNEN BEEINFLUSSEN DIE ABSATZZAHLEN NICHT

TESTVERTEILUNG: χ^2-VERTEILUNG

TESTNIVEAU: $\alpha = 1\%$

TESTSTATISTIK: $\chi^2_{EMP} = \sum \frac{(n_0 - n_e)^2}{n_e} = 60{,}75 = \chi^2_{EMP}$

n_0 n_e	0 BIS 30	31 BIS 60	61 BIS 90	
	$\frac{(30-40)^2}{40}$ +	$\frac{(30-50)^2}{50}$ +	$\frac{(60-30)^2}{30}$	120
	+ $\frac{(90-80)^2}{80}$ +	$\frac{(120-100)^2}{100}$ +	$\frac{(30-60)^2}{60}$	240
	120	150	90	360

$= 60{,}75 = \chi^2_{EMP}$

n_0 n_e	0 BIS 30	31 BIS 60	61 BIS 90	
	30 1/3 $\boxed{40}$	30 5/12 50	60 1/4 30	120
	90 1/3 80	120 5/12 100	30 1/4 60	240
	120	150	90	360

$$\boxed{\frac{\cancel{120}^{1/3}}{\cancel{360}} \cdot \frac{120}{\cancel{360}} \cdot \cancel{360} = \frac{1}{3} \cdot 120}$$

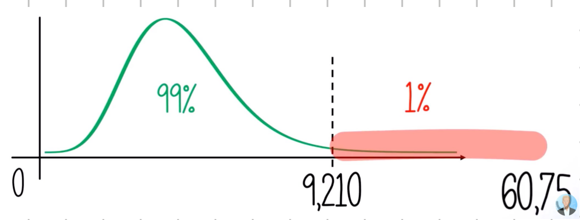
Alternative Berechnung:
(120x120)/360 = 40
(120x240)/360 = 80

2 FREIHEITSGRADE

30	30	60	120
90	120	30	240
120	150	90	360

Anhang B: Tabelle zur χ^2-Verteilung

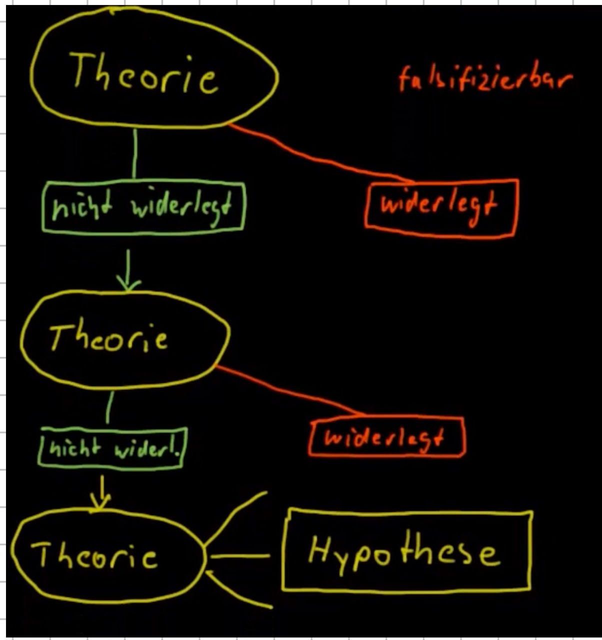
α / $1-\alpha$	0,9 / 0,1	0,75 / 0,25	0,5 / 0,5	0,25 / 0,75	0,1 / 0,9	0,05 / 0,95	0,025 / 0,975	0,01 / 0,99
1	0,016	0,102	0,455	1,323	2,706	3,841	5,024	6,635
2	0,211	0,575	1,386	2,773	4,605	5,991	7,378	$\boxed{9{,}210}$

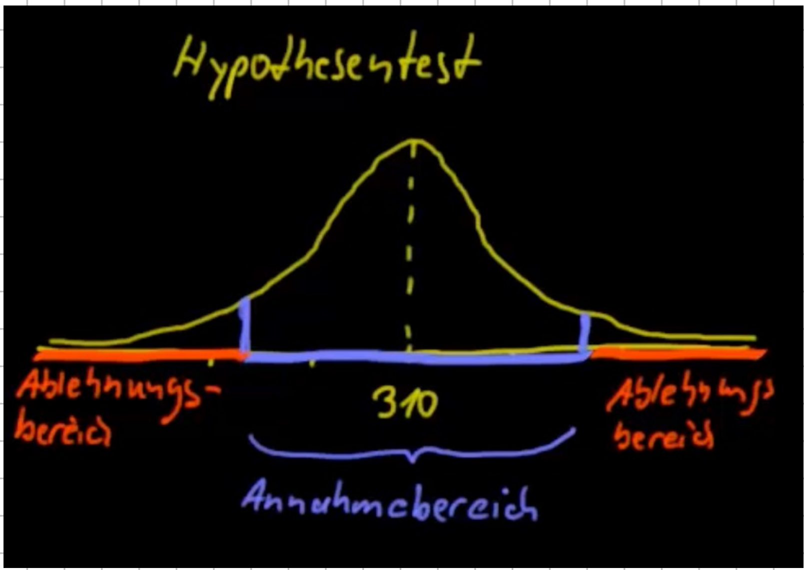


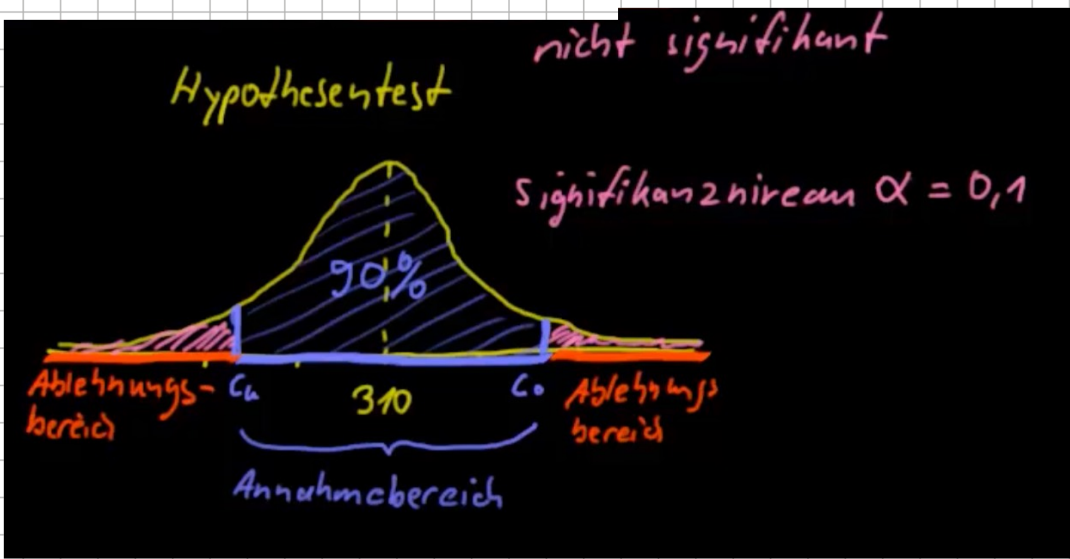
Anhang B: Tabelle zur χ^2-Verteilung

Freiheitsgrade	0,9	0,75	0,5	0,25	0,1	0,05	0,025	0,01
$1-\alpha$	0,1	0,25	0,5	0,75	0,9	0,95	0,975	0,99
1	0,016	0,102	0,455	1,323	2,706	3,841	5,024	6,635
2	0,211	0,575	1,386	2,773	4,605	5,991	7,378	9,210
3	0,584	1,213	2,366	4,108	6,251	7,815	9,348	11,345
4	1,064	1,923	3,357	5,385	7,779	9,488	11,143	13,277
5	1,610	2,675	4,351	6,626	9,236	11,070	12,833	15,086
6	2,204	3,455	5,348	7,841	10,645	12,592	14,449	16,812
7	2,833	4,255	6,346	9,037	12,017	14,067	16,013	18,475
8	3,490	5,071	7,344	10,219	13,362	15,507	17,535	20,090
9	4,168	5,899	8,343	11,389	14,684	16,919	19,023	21,666
10	4,865	6,737	9,342	12,549	15,987	18,307	20,483	23,209
11	5,578	7,584	10,341	13,701	17,275	19,675	21,920	24,725
12	6,304	8,438	11,340	14,845	18,549	21,026	23,337	26,217
13	7,042	9,299	12,340	15,984	19,812	22,362	24,736	27,688
14	7,790	10,165	13,339	17,117	21,064	23,685	26,119	29,141
15	8,547	11,037	14,339	18,245	22,307	24,996	27,488	30,578
16	9,312	11,912	15,338	19,369	23,542	26,296	28,845	32,000
17	10,085	12,792	16,338	20,489	24,769	27,587	30,191	33,409
18	10,865	13,675	17,338	21,605	25,989	28,869	31,526	34,805
19	11,651	14,562	18,338	22,718	27,204	30,144	32,852	36,191
20	12,443	15,452	19,337	23,828	28,412	31,410	34,170	37,566
21	13,240	16,344	20,337	24,935	29,615	32,671	35,479	38,932
22	14,041	17,240	21,337	26,039	30,813	33,924	36,781	40,289
23	14,848	18,137	22,337	27,141	32,007	35,172	38,076	41,638
24	15,659	19,037	23,337	28,241	33,196	36,415	39,364	42,980
25	16,473	19,939	24,337	29,339	34,382	37,652	40,646	44,314
26	17,292	20,843	25,336	30,435	35,563	38,885	41,923	45,642
27	18,114	21,749	26,336	31,528	36,741	40,113	43,195	46,963
28	18,939	22,657	27,336	32,620	37,916	41,337	44,461	48,278
29	19,768	23,567	28,336	33,711	39,087	42,557	45,722	49,588
30	20,599	24,478	29,336	34,800	40,256	43,773	46,979	50,892
31	21,434	25,390	30,336	35,887	41,422	44,985	48,232	52,191
32	22,271	26,304	31,336	36,973	42,585	46,194	49,480	53,486
33	23,110	27,219	32,336	38,058	43,745	47,400	50,725	54,776
34	23,952	28,136	33,336	39,141	44,903	48,602	51,966	56,061
35	24,797	29,054	34,336	40,223	46,059	49,802	53,203	57,342
36	25,643	29,973	35,336	41,304	47,212	50,998	54,437	58,619
37	26,492	30,893	36,336	42,383	48,363	52,192	55,668	59,893
38	27,343	31,815	37,335	43,462	49,513	53,384	56,896	61,162
39	28,196	32,737	38,335	44,539	50,660	54,572	58,120	62,428
40	29,051	33,660	39,335	45,616	51,805	55,758	59,342	63,691

Hypothesentests







Nullhypothese H_0: $\mu = 310$

Alternativhypothese H_1: $\mu \neq 310$

H_0 kann nur abgelehnt oder angenommen werden

α-Fehler: (Fehler 1. Art) Ablehnung einer richtigen H_0

β-Fehler: (Fehler 2. Art) Annahme einer falschen H_0

Kovarianz

Definition 3.2: Empirische Kovarianz

Die empirische Kovarianz wird definiert durch

$$s_{XY} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) \left(= \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \overline{x}\cdot\overline{y} \right).$$

Hier bezeichnen wie üblich $\overline{x}$ und $\overline{y}$ die arithmetischen Mittel der Merkmale X bzw. Y.

Beispiel 1:

i	1	2	3	4	5	6
Kugelstoßen x_i	$2^{-1,5}$ 8,06	$6^{2,5}$ 7,75	$1^{-2,5}$ 8,10	$4^{0,5}$ 7,9	$3^{-0,5}$ 8,01	$5^{1,5}$ 7,81
Weitsprung y_i	$4^{0,5}$ 3,98	$2^{-1,5}$ 4,15	$3^{-0,5}$ 4,07	$1^{-2,5}$ 4,18	$6^{2,5}$ 3,87	$5^{1,5}$ 3,95

$(x_i - \bar{x})$
$(y_i - \bar{y})$

$1 \Rightarrow 1 - 3,5 = -2,5 \quad | \quad 2 \Rightarrow 2 - 3,5 = -1,5$
$3 \Rightarrow 3 - 3,5 = -0,5 \quad | \quad 4 \Rightarrow 4 - 3,5 = 0,5$
$5 \Rightarrow 5 - 3,5 = 1,5 \quad | \quad 6 \Rightarrow 6 - 3,5 = 2,5$

$$\text{Cov}(x,y) = \frac{1}{n}\sum (x_i - \bar{x})\cdot(y_i - \bar{y})$$

$$\text{Cov}(x,y) = \frac{(-1,5)\cdot 0,5 + 2,5\cdot(-1,5) + (-2,5)\cdot(-0,5) + 0,5\cdot(-2,5) + (-0,5)\cdot 2,5 + 1,5\cdot 1,5}{6} = -\frac{7}{12}$$

Beispiel 2:

Student	Körpergröße X	Schuhgröße Y	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})\cdot(y_i - \bar{y})$
1	180	45	-4	0,5	-2
2	176	42	-8	-2,5	20
3	185	44	1	-0,5	-0.5
4	195	47	11	2,5	27,5

$\text{Var}(x) = s_x^2$; $\text{Var}(y) = s_y^2$; $\text{Cov}(x,y)$

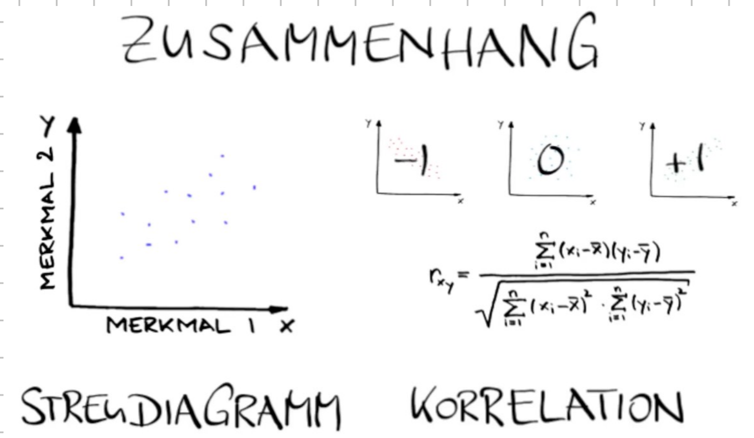
Kürzerer alternativer Weg!

x_i	y_i	$x_i \cdot y_i$
15	10	150
30	8	240
60	5	300
120	3	360
180	1	180
240	0.5	120
$\overline{x} = 107.5$	$\overline{y} = 4.583$	$\sum: 1350$

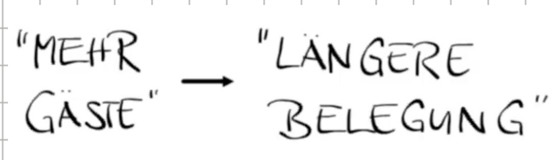
$$s_{XY} = \frac{1}{6}\cdot 1350 - 107.5\cdot 4.583 \approx -267.7.$$

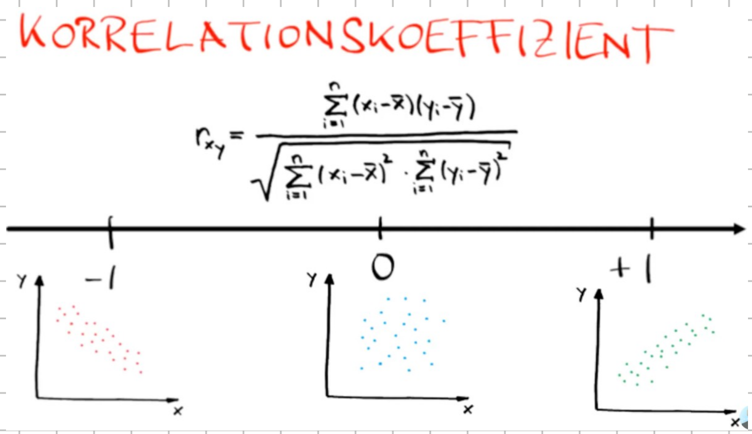
Streudiagramm und Korrelation

“Stärke des Zusammenhalts”

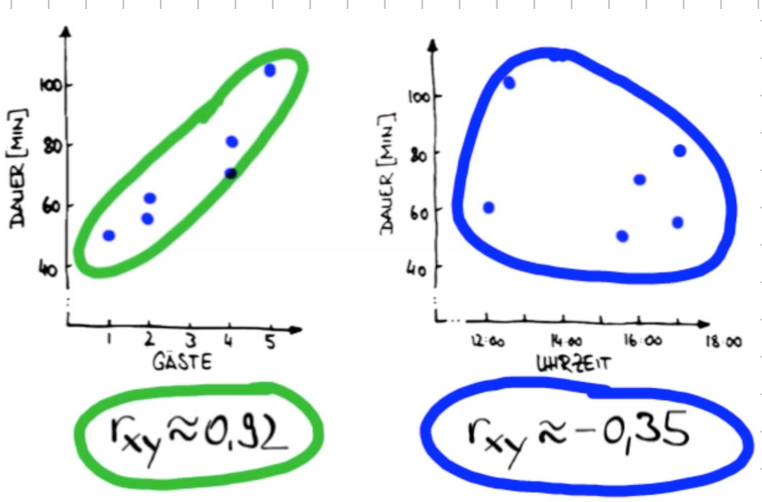


Wielange bleiben Personengruppen sitzen?





Werte ab $|r_{xy}| \geq 0.5$ deuten auf eine moderate Korrelation hin



Kausalität

Es könnte sein, dass zwei scheinbar abhängige Merkmale durch das ausblenden aller restlichen Merkmale zustande kommen. Daher ist es zwingend notwendig bei vermeindlich zusammenhängenden Merkmalen, weitere nachforschungen anzustellen um deren Abhängigkeit zu beweisen.

Korrelation und Korrelationskoeffizient

Vorgehen:

- Tabelle aufstellen
- Arithmetischen-Mittelwert x und y bestimmen
- Varianz x und y bestimmen
- Kovarianz x,y bestimme
- Korrelationskoeffizient bestimmen

Student	Körpergröße X	Schuhgröße Y
1	180	45
2	176	42
3	185	44
4	195	47

Bravais-Pearson

Arithmetisches Mittel von x und y bestimmen

$$\bar{x} = \frac{180+176+185+195}{4} = 184$$

$$\bar{y} = \frac{45+42+44+47}{4} = 44,5$$

Tabelle mit ergänzen

Student	Körpergröße X	Schuhgröße Y	$x_i - \bar{x}$	$y_i - \bar{y}$
1	180	45	-4	0,5
2	176	42	-8	-2,5
3	185	44	1	-0,5
4	195	47	11	2,5

Varianz s² für x und y bestimmen

$$s_x^2 = \frac{16+64+1+121}{4} = 50,5$$

$$s_y^2 = \frac{0,25+6,25+0,25+6,25}{4} = 3,25$$

Tabelle ergänzen

Student	Körpergröße X	Schuhgröße Y	$Var(x) = s_x^2$ $x_i - \bar{x}$	$Var(y) = s_y^2$ $y_i - \bar{y}$	$cov(x,y)$ $(x_i-\bar{x})\cdot(y_i-\bar{y})$
1	180	45	-4	0,5	-2
2	176	42	-8	-2,5	20
3	185	44	1	-0,5	-0.5
4	195	47	11	2,5	27,5

Kovarianz bestimmen

$$cov(x,y) = \frac{-2+20+(-0,5)+27,5}{4} = \frac{45}{4} = 11,25$$

Korrelationskoeffizient bestimmen

$$r = \frac{cov(x,y)}{s_x \cdot s_y} = \frac{11,25}{\sqrt{50,5 \cdot 3,25}}$$

s_x / s_y = Standardabweichung $= \sqrt{Var(x)} / \sqrt{Var(y)}$

Spearman-Pearson

i	1	2	3	4	5	6
Kugelstoßen x_i	8,06	7,75	8,10	7,9	8,01	7,81
Weitsprung y_i	3,98	4,15	4,07	4,18	3,87	3,95

Vorgehen:
- Rang bestimmen für Merkmale x und y
- Arithmetischer-Mittelwert bestimmen von Rang
- Varianzen von Rang x und Rang y bestimmen
- Kovarianz x,y bestimmen
- Rangkorrelationskoeffizient bestimmen

Rang bestimmen

i	1	2	3	4	5	6
Kugelstoßen x_i	2 8,06	6 7,75	1 8,10	4 7,9	3 8,01	5 7,81
Weitsprung y_i	4 3,98	2 4,15	3 4,07	1 4,18	6 3,87	5 3,95

Arithmetisches-Mittel x und y vom Rang

$$\bar{x} = \frac{2+6+1+4+3+5}{6} = \frac{21}{6} = 3,5$$

$$\bar{y} = \frac{21}{6} = 3,5$$

x und y enthalten die selben Zahlen!

Varianzen von Rang x und Rang y bestimmen

$$s_x^2 = \frac{(-1,5)^2 + 2,5^2 + (-2,5)^2 + 0,5^2 + (-0,5)^2 + 1,5^2}{6} = \frac{35}{12}$$

$$s_y^2 = \frac{35}{12}$$

x und y enthalten die selben Zahlen!

Kovarianz x,y, bestimmen

i	1	2	3	4	5	6
Kugelstoßen x_i	2 $^{-1,5}$ 8,06	6 ^{2,5} 7,75	1 $^{-2,5}$ 8,10	4 ^{0,5} 7,9	3 $^{-0,5}$ 8,01	5 ^{1,5} 7,81
Weitsprung y_i	4 ^{0,5} 3,98	2 $^{-1,5}$ 4,15	3 $^{-0,5}$ 4,07	1 $^{-2,5}$ 4,18	6 ^{2,5} 3,87	5 ^{1,5} 3,95

$(x_i - \bar{x})$
$(y_i - \bar{y})$

$1 \Rightarrow 1 - 3,5 = -2,5$ | $2 \Rightarrow 2 - 3,5 = -1,5$
$3 \Rightarrow 3 - 3,5 = -0,5$ | $4 \Rightarrow 4 - 3,5 = 0,5$
$5 \Rightarrow 5 - 3,5 = 1,5$ | $6 \Rightarrow 6 - 3,5 = 2,5$

$$Cov(x,y) = \frac{1}{n} \sum (x_i - \bar{x}) \cdot (y_i - \bar{y})$$

$$Cov(x,y) = \frac{(-1,5)\cdot 0,5 + 2,5 \cdot (-1,5) + (-2,5)\cdot(-0,5) + 0,5\cdot(-2,5) + (-0,5)\cdot 2,5 + 1,5 \cdot 1,5}{6} = -\frac{7}{12}$$

Rangkorrelationskoeffizient bestimmen

$$r_s = \frac{Cov(x,y)}{\sqrt{s_x^2 \cdot s_y^2}} = \frac{-\frac{7}{12}}{\frac{35}{12}} = -0,2$$

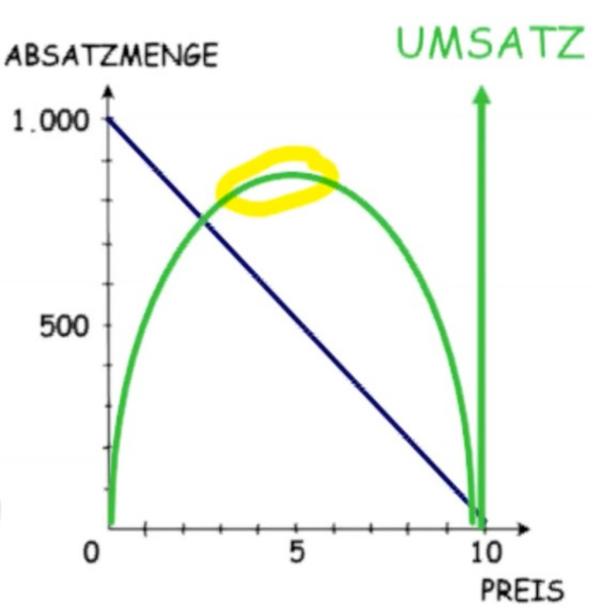
Kurzform wenn alle Ränge nur einmal vorkommen!

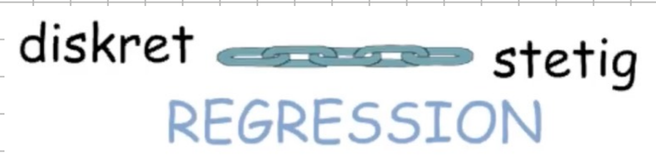
i	1	2	3	4	5	6	
Kugelstoßen x_i	2 8,06	6 7,75	1 8,10	4 7,9	3 8,01	5 7,81	
Weitsprung y_i	4 3,98	2 4,15	3 4,07	1 4,18	6 3,87	5 3,95	
d_i	2	4	2	3	3	0	Σ
d_i^2	4	16	4	9	9	0	42

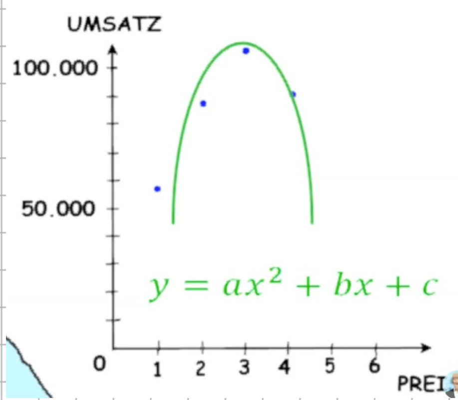
$$r_s = 1 - \frac{6 \cdot \sum d_i^2}{n \cdot (n^2 - 1)} = 1 - \frac{6 \cdot 42}{6 \cdot 35} = -0,2$$

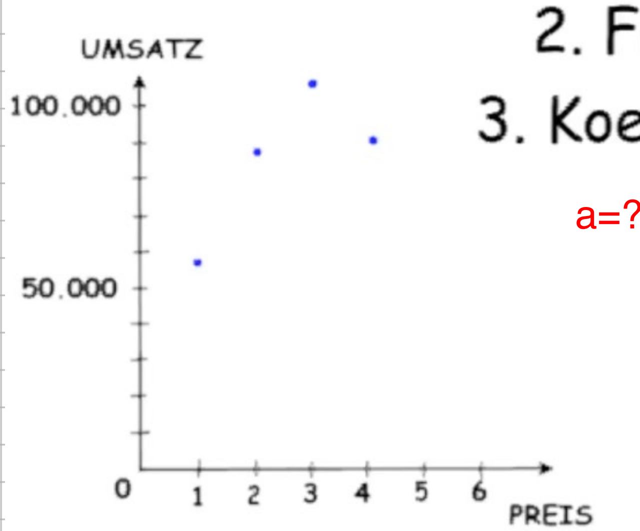
Regression

"Art des Zusammenhalts"



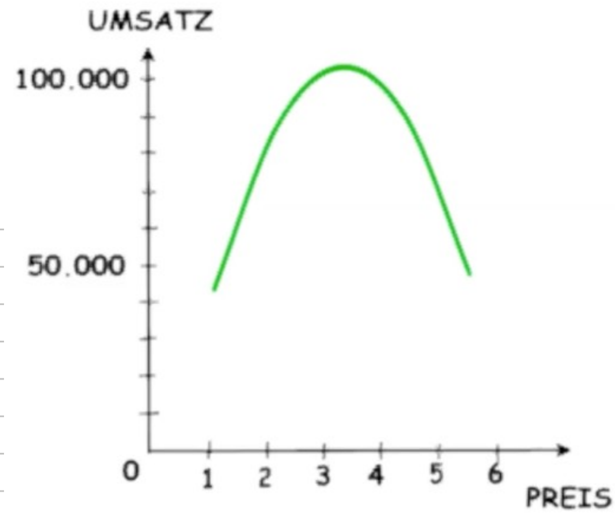






1. Messwerte
2. Funktionstyp festlegen
3. Koeffizienten bestimmen

a=? b=? c=?



METHODE DER KLEINSTEN FEHLERQUADRATE

$$\sum_{i=1}^{n}\left[y_i - f(x_i)\right]^2 \rightarrow MIN!$$

METHODE DER KLEINSTEN FEHLERQUADRATE

1. FEHLERFUNKTION AUFSTELLEN
2. ABLEITEN UND NULL SETZEN
3. UMSTELLEN, SUMMEN BERECHNEN UND EINSETZEN
4. GLEICHUNGSSYSTEM LÖSEN

ABWEICHUNG ZWISCHEN MESSWERT UND FUNKTIONSWERT: "MÖGLICHST KLEIN"

Abweichung möglichst nah bei “0”

$$E(a;b) = \sum_{i=1}^{3} (y_i - a - bx_i)^2$$

DIE ABLEITUNG EINER SUMME IST DIE SUMME DER ABLEITUNGEN!

$$E(a;b) = \sum_{i=1}^{3} (y_i - a - bx_i)^2$$

"MÖGLICHST KLEIN": MINIMUM

ABLEITEN

NULL SETZEN

$x_1 = 1 \quad x_2 = 3 \quad x_3 = 4$

$$E(a) = \sum_{i=1}^{3} x_i a = 1a + 3a + 4a$$

$$E'(a) = \sum_{i=1}^{3} x_i = 1 + 3 + 4$$

Summen können “ignoriert” werden →

$$E(a) = \sum_{i=1}^{3} x_i a$$

$$E'(a) = \sum_{i=1}^{3} x_i$$

$$E'_a(a;b) = \sum 2(y_i - a - bx_i)(-1)$$

$$E'_b(a;b) = \sum 2(y_i - a - bx_i)(-x_i)$$

$$\sum 2(y_i - a - bx_i)(-1) = 0$$

$$\sum 2(y_i - a - bx_i)(-x_i) = 0$$

$$2\sum (y_i - a - bx_i)(-1) = 0 \quad |:2$$

$$2\sum (y_i - a - bx_i)(-x_i) = 0 \quad |:2$$

$$\sum (y_i - a - bx_i)(-1) = 0$$

$$\sum (y_i - a - bx_i)(-x_i) = 0$$

$$\sum -y_i + a + bx_i = 0$$

$$\sum -x_iy_i + ax_i + bx_i^2 = 0$$

$$-\sum y_i + \sum a + \sum bx_i = 0$$

$$-\sum x_iy_i + \sum ax_i + \sum bx_i^2 = 0$$

$$\sum a + \sum bx_i = \sum y_i$$

$$\sum ax_i + \sum bx_i^2 = \sum x_iy_i$$

$$\sum a + b\sum x_i = \sum y_i$$

$$a\sum x_i + b\sum x_i^2 = \sum x_iy_i$$

$$a\,3 + b\sum x_i = \sum y_i$$

$$a\sum x_i + b\sum x_i^2 = \sum x_iy_i$$

DAUER KAFFEE

i	x_i	y_i
1	1	1,5
2	3	2
3	4	4
Σ	8	

$$a\,3 + b\,8 = \sum y_i$$

$$a\,8 + b\sum x_i^2 = \sum x_iy_i$$

DAUER KAFFEE

i	x_i	y_i
1	1	1,5
2	3	2
3	4	4
Σ	8	7,5

$$a\,3 + b\,8 = 7{,}5$$

$$a\,8 + b\sum x_i^2 = \sum x_iy_i$$

DAUER KAFFEE

i	x_i	y_i	x_i^2
1	1	1,5	1
2	3	2	9
3	4	4	16
Σ	8	7,5	26

$$a\,3 + b\,8 = 7{,}5$$

$$a\,8 + b\,26 = \sum x_iy_i$$

DAUER KAFFEE

i	x_i	y_i	x_i^2	x_iy_i
1	1	1,5	1	1,5
2	3	2	9	6
3	4	4	16	16
Σ	8	7,5	26	23,5

$$E(a;b) = \sum_{i=1}^{3} (y_i - a - bx_i)^2$$

$$a\,3 + b\,8 = 7{,}5$$

$$a\,8 + b\,26 = 23{,}5$$

Gauss-Algorythmus

$$E(a;b) = \sum_{i=1}^{3} (y_i - a - bx_i)^2$$

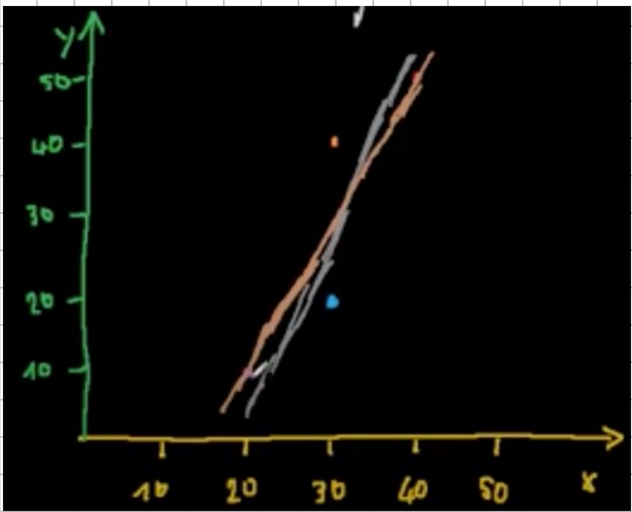
$$a = 0{,}5$$

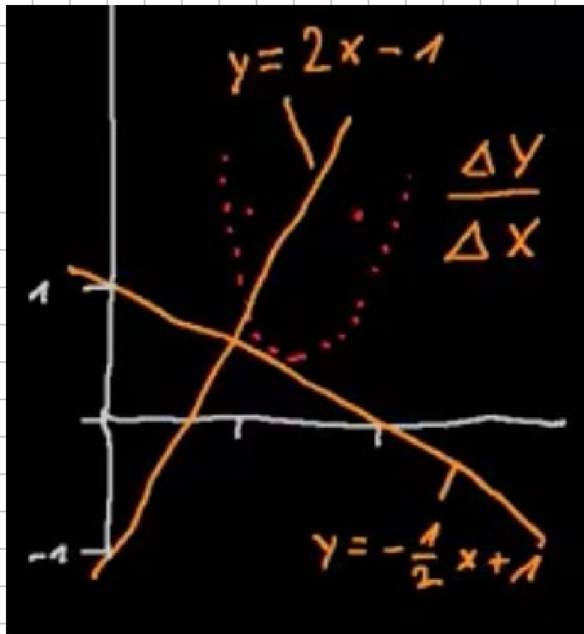
$$b = 0{,}75$$

Regressionsvariablen bestimmen

$$b = r_{XY}\frac{s_Y}{s_X} = \frac{s_{XY}}{s_X^2} \quad \text{und} \quad a = \overline{y} - b\overline{x}\,.$$

Student		1	2	3	4
Mathepunkte	x	30	40	20	30
Statistikpunkte	y	40	50	10	20





$$b = \frac{Cov(x,y)}{\tilde{s}_x \cdot \tilde{s}_x} = \frac{Cov(x,y)}{\tilde{s}_x^2}$$

$$b = \frac{100}{50} = 2$$

$$\hat{y} = \underline{b}x + \underline{a}$$

$$\sum (y_i - \hat{y}_i)^2 \text{ ist minimal}$$

$$\sum (y_i - (bx_i + a))^2$$

$$\overline{y} = b\overline{x} + a$$

$$a = \overline{y} - b\overline{x} = 30 - 2 \cdot 30 = -30$$

$$\boxed{\hat{y} = 2x - 30}$$

Aufgaben:

Gegeben sind die folgenden Werte für die Merkmale X und Y:

i	1	2	3	4	5	6
x_i	4	7	c	1	4	2
y_i	15	18	8	14	17	12

Berechnet wurde die Regressionsgerade $\hat{y} = 1{,}25x + 10{,}25$.
Bestimme den fehlenden Wert c in der Tabelle.

$$\bar{y} = \frac{15+18+8+14+17+12}{6} = 14$$

$$\hat{y} = bx + a$$

$$\bar{y} = b\bar{x} + a$$

$$\bar{y} = 1{,}25\,\bar{x} + 10{,}25$$

$$\Rightarrow 1{,}25\,\bar{x} = \bar{y} - 10{,}25$$

$$\Rightarrow \bar{x} = \frac{\bar{y} - 10{,}25}{1{,}25}$$

$$\bar{x} = \frac{14 - 10{,}25}{1{,}25} = 3$$

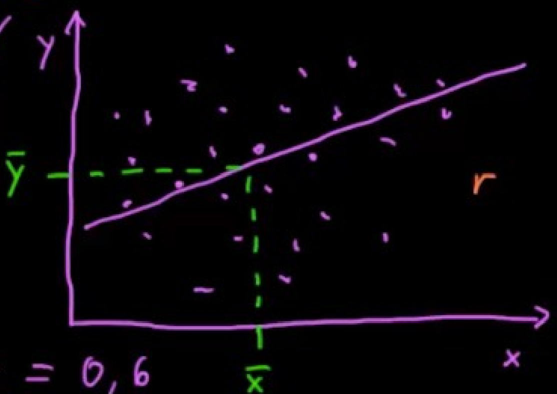
$$\Rightarrow c = 0$$

Gegeben ist der Parameter $a = 1$ der Regressionsfunktion $\hat{y} = a + bx$, sowie die folgenden Größen

$$\sum_{i=1}^{n} y_i = 400, \quad n = 100, \quad \sum_{i=1}^{n} {x_i}^2 = 2800, \quad \tilde{S}_X^2 = 3$$

Berechne den Parameter b, wobei $\bar{x} > 0$ gilt.

$$\bar{y} = b\bar{x} + a$$



$$b\bar{x} = \bar{y} - a$$

$$\Rightarrow b = \frac{\bar{y} - a}{\bar{x}} = \frac{3}{5} = 0{,}6$$

$$b = \frac{Cov(x,y)}{\tilde{S}_x^2}$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{400}{100} = 4$$

$$\tilde{S}^2 = \frac{1}{n}\sum x_i^2 - \bar{x}^2 \Rightarrow \bar{x}^2 = \frac{1}{n}\sum x_i^2 - \tilde{S}_x^2 = 28 - 3 = 25$$

$$\Rightarrow \bar{x} = 5$$