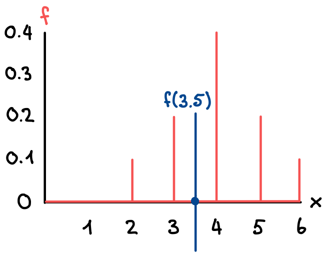
Beschreibende Statistik mit einem Merkmal

Häufigkeiten

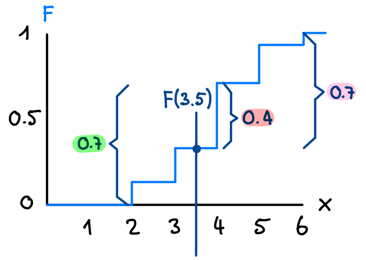
Funktion	z.B. Ganze Noten 1-6 (n=10)					
Merkmalsausprägung: a_k	1	2	3	4	5	6
absolute Häufigkeit: h_k	0	1	2	4	2	1
kumulative absolute Häufigkeit: $H_k = h_1 + \ldots + h_k$	0	1	3	7	9	10
relative Häufigkeit: $f_k = \frac{h_k}{n}$	0	1/10	2/10	4/10	2/10	1/10
kumulative relative Häufigkeit: $F_k = f_1 + \ldots + f_k$	0	1/10	3/10	7/10	9/10	1

Relative Häufigkeitsfunktion (PMF)



relative Häufigkeit $x = 3.5 = 0$

Kumulative Verteilungsfunktion (CDF)



Summe relativen Häufigkeiten $x \leq 3.5 = 0.3$

z.B. n = 10 mit Hilfe von CDF

Anzahl Daten genau (=) Wert 4 : $10 \cdot 0.4 = 4$

Anzahl Daten max. (≤) Wert 4 : $10 \cdot 0.7 = 7$

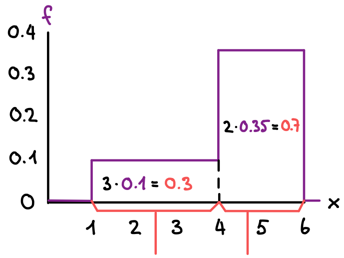
Anzahl Daten mind. (≥) Wert 4 : $10 \cdot 0.7 = 7$

Klassen

Zusammenfassung von mehreren Merkmalsausprägungen

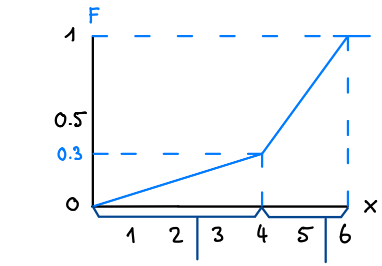
Funktion	z.B. Ganze Noten 1-6	
Klassenbereich : a_k bis a_{k+x}	[1, 4[1,2,3	[4, 6] 4,5,6
Klassenbreite: $b_k = a_{k+x} - a_k$	3	2
absolute Häufigkeit: h_k	3	7
relative Häufigkeit: $p_k = \frac{h_k}{n}$	3/10	7/10
relative Häufigkeitsdichte: $f_k = \frac{p_k}{b_k}$	$0.3/3 = 0.1$	$0.7/2 = 0.35$
kumulative relative Häufigkeit: $F_k = f_1 + \ldots + f_k$	3/10	1

Relative Häufigkeitsdichtefunktion (PDF)



Fläche = relative Häufigkeit

Kumulative Verteilungsfunktion (CDF)

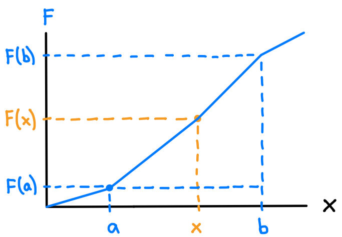


linearer Anstieg in Klassen

Lineare Interpolation / Quantile bei klassierten Daten

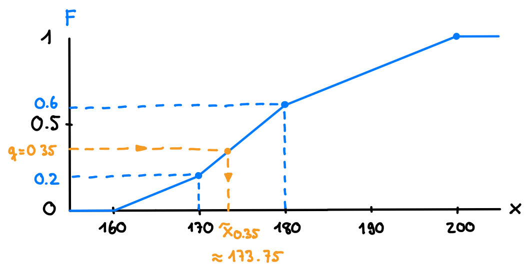
Bestimmung $\tilde{x}_q$ / $F(x)$ aus klassierter CDF

$$F(x) - F(a) = \frac{F(b) - F(a)}{b - a} \cdot (x - a)$$



z.B.

Klasse	[160, 170[	[170, 180[	[180, 200]
p_k	0.2	0.4	0.4
F_k	0.2	0.6	1



x für $F(x) = \tilde{x}_{0.35}$

$a = 170,\ b = 180,\ F(a) = 0.2,\ F(b) = 0.6$

$$\underbrace{0.35 - 0.2}_{0.15} = \frac{0.6 - 0.2}{\underbrace{180 - 170}_{0.04}} \cdot (\tilde{x}_{0.35} - 170)$$

$$0.15 = 0.04 \cdot \tilde{x}_{0.35} - 6.8$$

$$6.95 = 0.04 \cdot \tilde{x}_{0.35}$$

$$173.75 = \tilde{x}_{0.35}$$

x für $F(x) = \tilde{x}_{0.1}$

$a = 160,\ b = 170,\ F(a) = 0,\ F(b) = 0.2$

$$0.1 - 0 = \underbrace{\frac{0.2 - 0}{170 - 160}}_{0.02} \cdot (\tilde{x}_{0.1} - 160) \longrightarrow 0.1 = 0.02\,\tilde{x}_{0.1} - 3.2$$

$$165 = \tilde{x}_{0.1}$$

$F(x)$ für $x = 173.75$

$a = 170,\ b = 180,\ F(a) = 0.2,\ F(b) = 0.6$

$$F(173.75) - 0.2 = \frac{0.6 - 0.2}{180 - 170} \cdot (173.75 - 170)$$

$$F(173.75) = 0.15 + 0.2 = 0.35$$

Quantile bei nicht klassierten Daten

Gegeben: Aufsteigend sortierte Datenwerte $x_1 \leqslant \dots \leqslant x_n$, Anteil $0 < q < 1$

$$\tilde{x}_q = \begin{cases} \frac{1}{2}\cdot(x_{q\cdot n} + x_{q\cdot n+1}) & \text{wenn } q\cdot n \text{ ganzzahlig} \\ x_{\lceil q\cdot n \rceil} & \text{wenn } q\cdot n \text{ nicht ganzzahlig} \end{cases}$$

aufrunden auf nächste ganze Zahl

z.B. Körpergrössen Männer ($n = 10$): 164, 166, 174, 174, 174, 178, 182, 182, 188, 198

0.2-Quantil: $q\cdot n = 0.2 \cdot 10 = 2 \rightarrow$ ganzzahlig

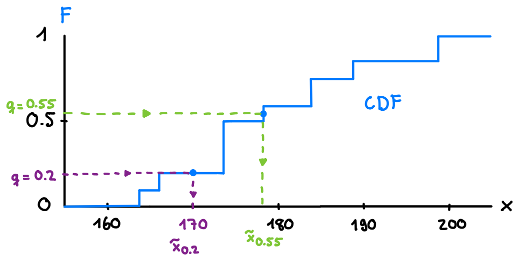
$\tilde{x}_{0.2} = \frac{1}{2}\cdot(x_2 + x_3) = \frac{1}{2}(166 + 174) = 170$

Anteil Daten unterhalb: 0.2 / oberhalb: $1 - 0.2 = 0.8$

0.55-Quantil: $q\cdot n = 0.55 \cdot 10 = 5.5 \rightarrow$ nicht ganzzahlig

$\tilde{x}_{0.55} = x_{\lceil 0.55 \rceil} = x_6 = 178$

Anteil Daten unterhalb: 0.55 / oberhalb: $1 - 0.55 = 0.45$



▬ : q-Wert auf Horizontalabschnitt, $\tilde{x}_q$ = mittlerer x-Wert

▬ : q-Wert auf Vertikalabschnitt, $\tilde{x}_q$ = x-Wert

Lagemasse

Charakterisiert das Zentrum der Daten, n Datenwerte $x_1, \dots, x_n$

Median: $\tilde{x}$ = 0.5-Quantil $\tilde{x}_{0.5}$

Arithmetisches Mittel / Mittelwert: $\bar{x} = \frac{x_1 + \dots + x_n}{n}$

↳ mit relativen Häufigkeiten $f_1, \dots, f_n$: $\bar{x} = (f_1 \cdot x_1) + \dots + (f_n \cdot x_n)$

z.B. Körpergrössen Männer

Median: $\tilde{x} = 176$

Mittelwert: $\bar{x} = \frac{164 + 166 + (3\cdot 174) + 178 + (2\cdot 182) + 188 + 198}{10} = \frac{1780}{10} = 178$

z.B. Ganze Noten 1-6 mit relativen Häufigkeiten

Mittelwert: $\bar{x} = (0\cdot 1) + (0.1\cdot 2) + (0.2\cdot 3) + (0.4 \cdot 4) + (0.2\cdot 5) + (0.1\cdot 6) = 4$

↳ Klassen (nur Approximation möglich):

Klassenmitten: $[1,4[\rightarrow 2.5$, $[4,6] \rightarrow 5$ $(6-4=2 \rightarrow 2:2=1 \rightarrow 6-1=5)$

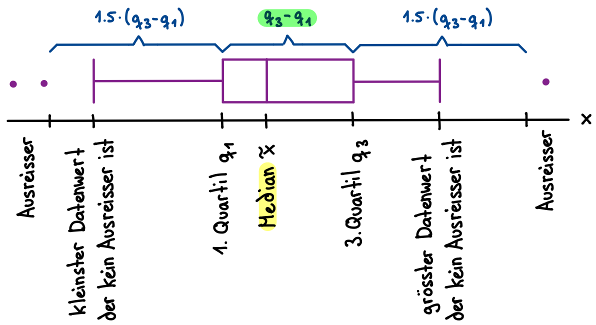
Mittelwert: $\bar{x} = (0.3\cdot 2.5) + (0.7 \cdot 5) = 4.25$

Boxplot

① Nur für nicht klassierte Daten

Aufsteigend sortiert

1. Quartil q_1 = 0.25-Quantil $\tilde{x}_{0.25}$
2. Quartil q_2 = Median $\tilde{x}$ = 0.5-Quantil $\tilde{x}_{0.5}$
3. Quartil q_3 = 0.75-Quantil $\tilde{x}_{0.75}$



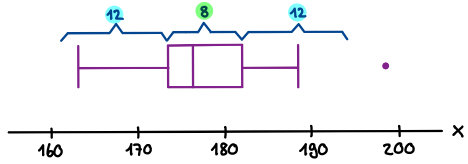
z.B. Körpergrössen Männer

1. Quartil $q_1 = 174$

Median $\tilde{x} = 176$

3. Quartil $q_3 = 182$

Interquartilsabstand $= 182 - 174 = 8 \rightarrow 1.5 \cdot 8 = 12$



Streuungsmasse

Charakterisiert Datenabweichung vom Zentrum, n Datenwerte $x_1, \dots, x_n$

Interquartilsabstand: 3. Quartil − 2. Quartil $= q_3 - q_2$

Varianz (Masseinheit im Quadrat): $\tilde{v} = \frac{(x_1 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{2} = \frac{x_1^2 + \dots + x_n^2}{n} - \bar{x}^2$

↳ mit relativen Häufigkeiten: $\tilde{v} = f_1 \cdot (x_1 - \bar{x})^2 + \dots + f_n \cdot (x_n - \bar{x})^2 = (f_1 \cdot x_1^2 + \dots + f_n \cdot x_n^2) - \bar{x}^2$

↳ korrigierte: $v = \frac{\tilde{v}\cdot n}{n-1}$

Resultat immer positiv oder null

Standardabweichung (Masseinheit identisch): $\tilde{s} = \sqrt{\tilde{v}}$

↳ korrigierte: $s = \sqrt{v}$

z.B. Körpergrössen Männer

Interquartilsabstand: $182 - 174 = 8$

Varianz: $\tilde{v} = \frac{164^2 + 166^2 + 3\cdot 174^2 + 178^2 + 2\cdot 182^2 + 188^2 + 198^2}{10} - 178^2 = 92$

↳ korrigierte: $v = \frac{92 \cdot 10}{10-1} = 102.\bar{2}$

Standardabweichung: $\tilde{s} = \sqrt{92} \approx 9.5$

↳ korrigierte: $s = \sqrt{102.\bar{2}} \approx 10.1$

z.B. Ganze Noten 1-6 mit relativen Häufigkeiten

Varianz: $\tilde{v} = (0\cdot 1^2 + 0.1 \cdot 2^2 + 0.2\cdot 3^2 + 0.4 \cdot 4^2 + 0.2 \cdot 5^2 + 0.1 \cdot 6^2) - 4^2 = 1.2$

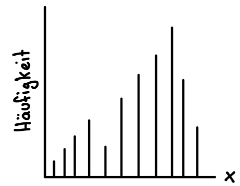
↳ Klassen (nur Approximation möglich):

Klassenmitten: $[1,4[\rightarrow 2.5$ $(4-1=3 \rightarrow 3:2=1.5 \rightarrow 4-1.5=2.5)$, $[4,6] \rightarrow 5$

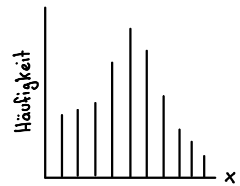
Varianz: $\tilde{v} = (0.3\cdot 2.5^2 + 0.7 \cdot 5^2) - 4^2 = 3.375$

Form der Datenverteilung

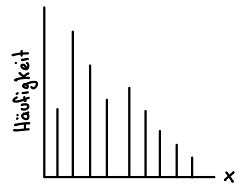
Mittelwert: $\bar{x}$, Median: $\tilde{x}$



$\bar{x} < \tilde{x}$: asymmetrisch linksschief



$\bar{x} \approx \tilde{x}$: näherungsweise symmetrisch

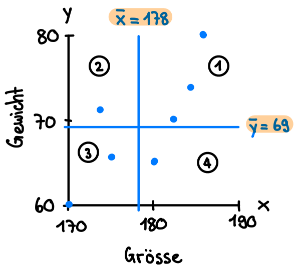


$\bar{x} > \tilde{x}$: asymmetrisch rechtsschief

Beschreibende Statistik mit mehreren Merkmalen

Streudiagramm

z.B. Zusammenhang Grösse und Gewicht									Summe	Mittelwert
Grösse in cm	186	174	170	174	180	175	182	183	1'424	178
Gewicht in kg	80	63	60	72	65	68	70	74	552	69



Quadrant Bereiche:

① Beide Merkmale überdurchschnittlich
② 1. Merkmal unter- / 2. überdurchschnittlich
③ Beide Merkmale unterdurchschnittlich
④ 1. Merkmal über- / 2. unterdurchschnittlich

Kovarianz

Abweichungsmittelwert der Datenwerte zum Datenmittelwert, n Datenpaare $(x_1, y_2), \dots, (x_n, y_2)$ mit Mittelwerten $\bar{x}, \bar{y}$

$$\tilde{v}_{xy} = \frac{(x_1-\bar{x})\cdot(y_1-\bar{y}) + \dots + (x_n-\bar{x})\cdot(y_n-\bar{y})}{n} = \frac{x_1\cdot y_1 + \dots + x_n\cdot y_n}{n} - \bar{x}\cdot\bar{y} \quad , \quad \tilde{v}_{xy} = \frac{rg(x_1)\cdot rg(y_1) + \dots + rg(x_n)\cdot rg(y_n)}{n} - \bar{x}\cdot\bar{y}$$

z.B. Zusammenhang Grösse und Gewicht (Datenwerte)

$$\tilde{v}_{xy} = \frac{186\cdot 80 + 174\cdot 63 + 170\cdot 60 + 174\cdot 72 + 180\cdot 65 + 175\cdot 68 + 182\cdot 70 + 183\cdot 74}{8} - 178\cdot 69 = 24.5$$

z.B. Zusammenhang Grösse und Gewicht (Datenränge)

$$\tilde{v}_{xy} = \frac{8\cdot 8 + 2.5\cdot 2 + 1\cdot 1 + 2.5\cdot 6 + 5\cdot 3 + 4\cdot 4 + 6\cdot 5 + 7\cdot 7}{8} - 4.5\cdot 4.5 = 4.125$$

Korrelation

Statistischer linearer Zusammenhang zweier Merkmale

Positive Korrelation: meisten Datenpunkte in ① und ③ → Kovarianz positiv

Negative Korrelation: meisten Datenpunkte in ② und ④ → Kovarianz negativ

Keine Korrelation: etwa gleich viele Datenpunkte in ① und ④ → Kovarianz nahe null

z.B. Grösse und Gewicht sind positiv korreliert, gross → gross / klein → klein

↳ wenn negativ korreliert, gross → klein / klein → gross

Datenränge anstatt Datenwerte

Minderung des Einflusses von Ausreissern

! Bei gleichem Wert geteilter Rang verwenden. z.B. Rang 2 und 3 gleich → $\frac{2+3}{2} = 2.5$

z.B. Zusammenhang Grösse und Gewicht									Summe	Mittelwert
rg(Grösse)	8	2.5	1	2.5	5	4	6	7	36	4.5
rg(Gewicht)	8	2	1	6	3	4	5	7	36	4.5

Korrelationskoeffizient

! Pearson: Datenwerte / Spearman: Datenränge

Standardabweichungen $\tilde{s}_x, \tilde{s}_y$

$r_{xy} = \frac{\tilde{v}_{xy}}{\tilde{s}_x \cdot \tilde{s}_y}$ (keine Masseinheit) } Resultat von -1 bis 1

Korrelation Faustregel:

-1 bis -0.75 / 0.75 bis 1	stark negativ / positiv
-0.75 bis -0.5 / 0.5 bis 0.75	moderat negativ / positiv
-0.5 bis -0.25 / 0.25 bis 0.5	schwach negativ / positiv
-0.25 bis 0.25	keine

z.B. Zusammenhang Grösse und Gewicht (Pearson)

$\tilde{s}_x = \sqrt{26.75}$, $\tilde{s}_y = \sqrt{36.25}$, $\tilde{v}_{xy} = 24.5$

$r_{xy} \approx \frac{24.5}{5.17\cdot 6.02} \approx 0.79$ → stark positive Korrelation

z.B. Zusammenhang Grösse und Gewicht (Spearman)

$\tilde{s}_x = \sqrt{5.18}$, $\tilde{s}_y = \sqrt{5.25}$, $\tilde{v}_{xy} = 4.125$

$r_{xy} \approx \frac{4.125}{2.27\cdot 2.29} \approx 0.79$ → stark positive Korrelation

Hilfstabellen Berechnung

Pearson:

i	...	Summe	Mittelwert
x_i	...		
y_i	...		
x_i^2	...		
y_i^2	...		
$x_i \cdot y_i$	...		

Spearman:

i	...	Summe	Mittelwert
x_i	...		
y_i	...		
$rg(x_i)$	...		
$rg(y_i)$	...		
$rg(x)^2$	...		
$rg(y)^2$	...		
$rg(x_i)\cdot rg(y_i)$	...		

Kombinatorik

Produktregel: Zählen geht über k Stufen
Insgesamt: $n_1 \cdot \ldots \cdot n_k$ Möglichkeiten

Summenregel: Beim Zählen treten j verschiedene Fälle auf
Insgesamt: $m_1 + \ldots + m_j$ Möglichkeiten

Komplementregel: Beim Zählen gibt es n Möglichkeiten
p Möglichkeiten mit der Eigenschaft $(p = n - q)$
q übrigen Möglichkeiten ohne die Eigenschaft

Beispiele

4-stellige PIN-Codes aus Ziffern 0 bis 9, Anzahl Kombinationen

mit lauter geraden Ziffern:
↳ $5^4 = 625$

mit lauter verschiedenen ungeraden Ziffern:
↳ $5 \cdot 4 \cdot 3 \cdot 2 = 120$

mit entweder lauter gleichen oder verschiedenen Ziffern:
1. Fall alle gleich: $10 \cdot 1 \cdot 1 \cdot 1 = 10$
2. Fall alle verschieden: $10 \cdot 9 \cdot 8 \cdot 7 = 5'040$

$10 + 5'040 = 5'050$

mit mind. 2 gleichen Ziffern:
n alle Möglichkeiten: $10^4 = 10'000$
q alle Ziffern verschieden: $10 \cdot 9 \cdot 8 \cdot 7 = 5'040$
$10'000 - 5'040 = 4'960$

mit max. 3 gleichen Ziffern:
n alle Möglichkeiten: $10^4 = 10'000$
q alle Ziffern gleich: $10 \cdot 1 \cdot 1 \cdot 1 = 10$
$10'000 - 10 = 9'990$

mit mind. je einer geraden und ungeraden Ziffer:
n alle Möglichkeiten: $10^4 = 10'000$
q alle gerade/ungerade: 1. Fall alle gerade: $5^4 = 625$
2. Fall alle ungerade: $5^4 = 625$
$625 + 625 = 1'250$ → $10'000 - 1'250 = 8'750$

mit genau einer Ziffer 9: $\binom{4}{1} \cdot 1 \cdot 9^3 = 2'916$

Typ 1: mit Reihenfolge mit Wiederholung

Möglichkeiten zur Auswahl von k aus n Objekten:

$n \cdot n \cdot \ldots \cdot n = n^k$

z.B. Tupel (X, Y, Z) mit X, Y, Z je A, B, C, D, E

$5 \cdot 5 \cdot 5 = 5^3 = 125$

Typ 2: mit Reihenfolge ohne Wiederholung

Möglichkeiten zur Auswahl von k aus n Objekten:

$n \cdot (n-1) \cdot \ldots \cdot (n-k+1) = \frac{n!}{(n-k)!}$ ⓘ $0! = 1$, PRB → !

z.B. Tupel (X, Y, Z) mit X, Y, Z je verschiedenen A, B, C, D, E

$5 \cdot 4 \cdot 3 = \frac{5!}{(5-3)!} = 60$

z.B. Wettbewerb mit 10 Personen, eine davon ist Alice

Möglichkeiten zur Belegung der ersten 3 Plätze:
↳ $10 \cdot 9 \cdot 8 = \frac{10!}{(10-3)!} = 720$

Möglichkeiten für Podest mit Alice:
1. Fall 1. Platz: $1 \cdot 9 \cdot 8 = 1 \cdot \frac{9!}{(9-2)!} = 72$
2. Fall 2. Platz: $9 \cdot 1 \cdot 8 = 1 \cdot \frac{9!}{(9-2)!} = 72$
3. Fall 3. Platz: $9 \cdot 8 \cdot 1 = 1 \cdot \frac{9!}{(9-2)!} = 72$

$72 + 72 + 72 = 216$

z.B. Kurs mit 20 Personen, 8 Frauen und 12 Männer

Möglichkeiten Einerreihe zu bilden:
↳ $\frac{20!}{(20-20)!} = 20! = 2.43 \cdot 10^{18}$

Möglichkeiten Einerreihe zu bilden mit Frauen vorne:
↳ $\frac{8!}{(8-8)!} \cdot \frac{12!}{(12-12)!} = 8! \cdot 12! = 1.93 \cdot 10^{13}$

Beispiel Kombination Typ 1, 2 und 3

Anzahl Wörter der Länge 5 mit 2 Vokalen und 3 verschiedenen Konstanten:

1. 2 von 5 Stellen an denen Vokale stehen:
↳ Typ 3: $\binom{5}{2} = 10$
2. Vokale an in 1. ermittelten Stellen:
↳ Typ 1: $2^2 = 4$
3. Verschiedene Konstanten an in 1. nicht ermittelten Stellen:
↳ Typ 2: $4 \cdot 3 \cdot 2 = \frac{4!}{1!} = 24$

→ Total: $10 \cdot 4 \cdot 24 = 960$

Typ 3: ohne Reihenfolge ohne Wiederholung

Möglichkeiten zur Auswahl von k aus n Objekten:

$\frac{n \cdot (n-1) \cdot \ldots \cdot (n-k+1)}{k \cdot (k-1) \cdot \ldots \cdot 2 \cdot 1} = \frac{n!}{k! \cdot (n-k)!} = \binom{n}{k}$ ⓘ PRB → nCr

z.B. Kurs mit 20 Personen, 8 Frauen und 12 Männer, Bildung einer 4er Gruppe

ⓘ Ohne Wiederholung: Anzahl Personen gegeben (Person A in nur 1er Gruppe)

Anzahl möglicher Gruppen insgesamt:
↳ $\binom{20}{4} = 4'845$

Anzahl Gruppen mit nur Männern oder nur Frauen:
1. Fall nur Männer: $\binom{12}{4} = 495$
2. Fall nur Frauen: $\binom{8}{4} = 70$

$495 + 70 = 565$

Anzahl Gruppen mit je 2 Männern und 2 Frauen:
↳ $\binom{12}{2} \cdot \binom{8}{2} = 1'848$

Anzahl Gruppen mit mind. 1 Mann und Frau:
1. Fall 2 Männer und 2 Frauen: 1'848
2. Fall 3 Männer und 1 Frau: $\binom{12}{3} \cdot \binom{8}{1} = 1'760$
3. Fall 1 Mann und 3 Frauen: $\binom{12}{1} \cdot \binom{8}{3} = 672$

$1'848 + 1'760 + 672 = 4'280$

Typ 4: ohne Reihenfolge mit Wiederholung

Möglichkeiten zur Auswahl von k aus n Objekten:

$\frac{(k+n-1)!}{k! \cdot (n-1)!} = \binom{k+n-1}{k}$ ⓘ PRB → nCr

z.B. Ausreichend viele Sugus mit und ohne Zucker in 4 Geschmacksrichtungen (8 Sorten)

ⓘ Mit Wiederholung: Anzahl Sugus nicht gegeben (Sorte mehrfach ziehbar)

Anzahl Kombinationen wenn 6 Sugus gezogen werden:
↳ $\binom{6+8-1}{6} = 1'716$

Anzahl Kombinationen wenn je 3 Sugus mit und ohne Zucker gezogen werden:
↳ $\binom{3+4-1}{3} \cdot \binom{3+4-1}{3} = 400$

Anzahl Kombinationen wenn 5 Sugus ohne Erdbeer gezogen werden und alle entweder mit oder ohne Zucker sind:
1. Fall mit Zucker: $\binom{5+3-1}{5} = 21$
2. Fall ohne Zucker: $\binom{5+3-1}{5} = 21$

$21 + 21 = 42$

Anzahl Kombinationen wenn 5 Sugus davon mind. 1 Erdbeer gezogen werden:
n alle Kombinationen: $\binom{5+8-1}{5} = 792$
q ohne Erdbeer: $\binom{5+6-1}{5} = 252$

$792 - 252 = 540$

Elementare Wahrscheinlichkeitsrechnung

Zufallsexperiment

Vorgang mit mehreren möglichen Ausgängen und einem ungewissen eintretenden Ausgang

Ergebnisse

Ergebnisse: Mögliche Ausgänge eines Zufallsexperiments
Bezeichnung mit $\omega_1, \omega_2, \dots$

Ergebnisraum: Menge aller Ergebnisse
Bezeichnung mit $\Omega = \{\omega_1, \omega_2, \dots\}$

Ereignisse

Ereignisse: Teilmengen des Ergebnisraums Ω
Bezeichnung mit $A \subseteq \Omega, B \subseteq \Omega, C \subseteq \Omega, \dots$

Elementarereignis: Ereignis mit genau einem Ergebnis

Zusammengesetztes Ereignis: Ereignis mit mehr als einem Ergebnis

Unmögliches Ereignis: Leere Menge $\{\}$, tritt nie ein

Sicheres Ereignis: Ganze Ergebnismenge Ω, tritt immer ein

Beispiel

Zufallsexperiment: $\underbrace{\text{2-Mal gleicher}}$ Würfel mit Augenzahlen 1,2,3 werfen
mit Reihenfolge

Ergebnisse: Augenzahlpaare $\omega_1 = (1,1), \omega_2 = (1,2), \dots, \omega_9 = (3,3)$

Ergebnisraum: $\Omega = \{\omega_1, \omega_2, \dots, \omega_9\} = \{(1,1), (1,2), \dots, (3,3)\}$

Ereignisse: A = Nur ungerade Augenzahlen $= \{(1,1),(1,3),(3,1),(3,3)\}$, $P(A) = 4/9$
↳ Zusammengesetztes Ereignis

B = Nur gerade Augenzahlen $= \{(2,2)\}$, $P(B) = 1/9$
↳ Elementarereignis

C = Augenzahlprodukt $5 = \{\}$, $P(C) = 0$
↳ Unmögliches Ereignis

D = Augenzahlunterschied $\leq 2 = \Omega$, $P(D) = 1$
↳ Sicheres Ereignis, Zusammengesetztes Ereignis

Wahrscheinlichkeit /-smass

Wahrscheinlichkeit von A: Zahl $P(A)$

Wahrscheinlichkeitsmass auf Ω: Funktion P

Eigenschaften:
- $0 \leq P(A) \leq 1$ für alle A
- $P(\{\}) = 0$ und $P(\Omega) = 1$
- $P(A \cup B \cup C \cup \dots) = P(A) + P(B) + P(C) + \dots$ für alle paarweise disjunkt
  $A, B, C, \dots$ mit $A \cap B = \{\}, A \cap C = \{\}, B \cap C = \{\}, \dots$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- $P(\bar{A}) = P(\Omega \setminus A) = 1 - P(A)$

z.B. $P(\bar{C}) = 0.7$, $A \cap C = \{\}$, $P(A \cup C) = 0.9$, $P(A \cap B) = 0.3$, $P(A \cup B) = 0.8$

$P(C) = 1 - P(\bar{C}) = 1 - 0.7 = 0.3$

$P(A) = P(A \cup C) - P(C) + P(A \cap C) = 0.9 - 0.3 + 0 = 0.6$

$P(B) = P(A \cup B) - P(A) + P(A \cap B) = 0.8 - 0.6 + 0.3 = 0.5$

Laplace-Zufallsexperiment

Alle elementaren Ereignisse haben dieselbe Wahrscheinlichkeit

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\text{Anzahl für A günstige Ergebnisse}}{\text{Anzahl insgesamt mögliche Ergebnisse}}$$

z.B. 10 Kugeln von 1-10 nummeriert, 4 Kugeln ohne zurücklegen $\underbrace{\text{nacheinander}}_{\text{mit Reihenfolge}}$ ziehen
↳ $|\Omega| = 10 \cdot 9 \cdot 8 \cdot 7 = 5'040$ (Typ 2)

Alle gezogene Kugeln gerade Nummer:
↳ $|A| = 5 \cdot 4 \cdot 3 \cdot 2 = 120$, $P(A) = \frac{120}{5'040} \approx 0.0238$

Erste beiden Kugeln gerade und letzten beiden ungerade Nummern:
↳ $|B| = 5 \cdot 4 \cdot 5 \cdot 4 = 400$, $P(B) = \frac{400}{5'040} \approx 0.0794$

Genau eine Kugel gerade Nummer hat:
↳ $|C| = \underbrace{\binom{4}{1}}_{\text{Fälle}} \cdot \underbrace{5}_{\text{gerade}} \cdot \underbrace{5 \cdot 4 \cdot 3}_{\text{ungerade}}$, $P(C) = \frac{1'200}{5'040} \approx 0.2381$

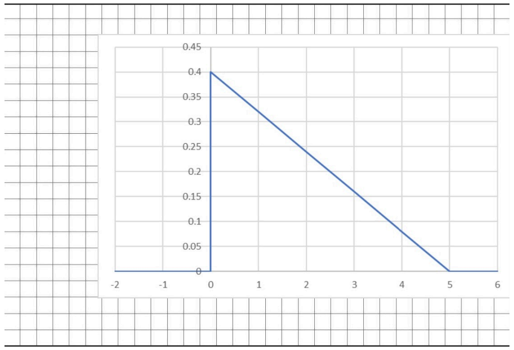
Mind. eine Kugel ungerade Nummer hat:
↳ $|\bar{D}| = |A| = 120 \longrightarrow P(\bar{D}) = P(A) \approx 0.0238$
$P(D) = 1 - P(\bar{D}) \approx \underline{\underline{0.9762}}$

Wahrscheinlichkeitsverteilung Dichtefunktion PDF

Kein Platz mehr bei Wahrscheinlichkeitsverteilung

$$f_X(t) = \begin{cases} (10-2t)/25 & \text{falls } 0 \leq t \leq 5 \\ 0 & \text{sonst} \end{cases}$$

(1P) Skizzieren Sie die Dichtefunktion.



(1P) Berechnen Sie die Wahrscheinlichkeit $P(X \geq 3)$.
$P(X \geq 3) = 4/25 = 0.16$

(4P) Berechnen Sie den Erwartungswert $E(X)$ sowie die Varianz $V(X)$.

$E(X) = \int_{-\infty}^{\infty} t \cdot f_X(t)\, dt = \int_0^5 1/25 \cdot (10t - 2t^2)\, dt = 1/25 \cdot \left[5t^2 - \frac{2}{3t}^3\right]_0^5 =$
$1/25 \cdot (125 - 250/3) = 5/3 = 1.\overline{6}$ und

$E(X^2) = \int_{-\infty}^{\infty} t^2 \cdot f_X(t)\, dt = \int_0^5 1/25 \cdot (10t^2 - 2t^3)\, dt = 1/25 \cdot \left[\frac{10}{3t^3} - \frac{2}{4t}^4\right]_0^5 =$
$1/25 \cdot (10/3 \cdot 125 - 1/2 \cdot 625) = 25/6$
$V(X) = 25/6 - (5/3)\text{^}2 = 25/18 = 1.3\overline{8}$

Bedingte Wahrscheinlichkeit

A tritt ein, wenn B eingetreten: $P(A|B) = \frac{P(A \cap B)}{P(B)}$

B tritt ein, wenn A eingetreten: $P(B|A) = \frac{P(B \cap A)}{P(A)}$

Multiplikationssatz: $P(A \cap B) = P(B \cap A) = P(A) \cdot P(B|A) = P(B) \cdot P(A|B)$

z.B. In Klasse mit 24-Schüler haben 18 genügende Note in Mathe und 20 genügende Note in Deutsch. 16 Schüler in beiden Fächer genügend.

Angaben in Vierfeldtafel darstellen:

	D	$\bar{D}$	
M	16	2	18
$\bar{M}$	4	2	6
	20	4	24

: Gegeben

Wahrscheinlichkeit ein Schüler in beiden Fächern ungenügend:

↳ $P(\bar{M} \cap \bar{D}) = 2/24 = 1/12$

Wahrscheinlichkeit ein Schüler entweder in Mathe oder Deutsch ungenügend:

↳ $P(\bar{M} \cap D) + P(M \cap \bar{D}) = 4/24 + 2/24 = 6/24 = 1/4$

Wahrscheinlichkeit ein Schüler, der in Deutsch genügend ist, es auch in Mathe ist:

↳ $P(M|D) = P(M \cap D) : P(D) = 16/24 : 20/24 = 16/20 = 4/5$

Wahrscheinlichkeit ein Schüler, der in Mathe ungenügend ist, es auch in Deutsch ist:

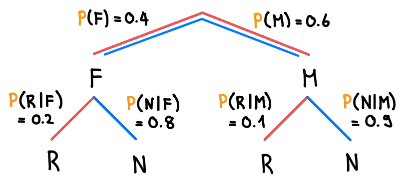
↳ $P(\bar{D}|\bar{M}) = P(\bar{D} \cap \bar{M}) : P(\bar{M}) = 2/24 : 6/24 = 2/6 = 1/3$

Satz von Bayes

Gegeben: $P(A)$, $P(B)$ und $P(B|A)$ / Gesucht: $P(A|B)$

$P(A|B) = \frac{P(A) \cdot P(B|A)}{P(B)}$ ⓘ Wechsel der Bedingung

z.B. Kurs mit 40% Frauen davon rauchen 20%, 60% Männer davon rauchen 10%



$P(F \cap R) + P(M \cap R) = 0.4 \cdot 0.2 + 0.6 \cdot 0.1 = 0.14$

$P(F \cap N) + P(M \cap N) = 0.4 \cdot 0.8 + 0.6 \cdot 0.9 = 0.86$

Zufällig ausgewählte rauchende Person ist ein Mann: $P(M|R) = \frac{P(M) \cdot P(R|M)}{P(R)} = \frac{0.6 \cdot 0.1}{0.14} \approx 0.429$

Zufällig ausgewählte nichtrauchende Person ist eine Frau: $P(F|N) = \frac{P(F) \cdot P(N|F)}{P(N)} = \frac{0.4 \cdot 0.8}{0.86} \approx 0.372$

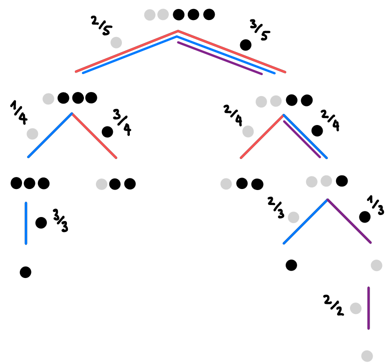
Ereignisbaum und Pfadregel

Situation mit bedingter Wahrscheinlichkeit anschaulich darstellen

Pfadwahrscheinlichkeit: Wahrscheinlichkeiten entlang des Pfades multiplizieren

Ereigniswahrscheinlichkeit: Summe aller Pfadwahrscheinlichkeit die zum Ergebnis führen

z.B. Gefäss mit 2 weissen und 3 schwarzen Kugeln. Solange eine Kugel ziehen und beiseite legen bis mind. 1 weisse und 1 schwarze draussen liegt.



Dazu 2-Mal ziehen:

↳ $2/5 \cdot 3/4 + 3/5 \cdot 2/4 = 0.6$

Dazu 3-Mal ziehen:

↳ $2/5 \cdot 1/4 \cdot 3/3 + 3/5 \cdot 2/4 \cdot 2/3 = 0.3$

Dazu 4-Mal ziehen:

↳ $3/5 \cdot 2/4 \cdot 1/3 \cdot 2/2 = 0.1$

Unabhängigkeit von Ereignissen

A und B sind unabhängig wenn gilt: $P(A|B) = P(A)$, $P(B|A) = P(B)$ oder $P(A \cap B) = P(A) \cdot P(B)$

z.B. Gewöhnlicher Würfel wird 1-Mal geworfen

$\Omega = \{1,2,3,4,5,6\}$, $A = \{2,4,6\}$, $B = \{1,2,3\}$, $C = \{3,4,5,6\}$

A und B: $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\{2\})}{1/2} = \frac{1/6}{1/2} = 1/3 \neq P(A) = 1/2 \longrightarrow$ abhängig

A und C: $P(A|C) = \frac{P(A \cap C)}{P(C)} = \frac{P(\{4,6\})}{2/3} = \frac{1/3}{2/3} = 1/2 = P(A) = 1/2 \longrightarrow$ unabhängig

↳ $P(A \cap C) = 1/2 \cdot 2/3 = 1/3$

z.B. Vierfeldtafel Mathe und Deutsch Noten Beispiel

M und D: $P(M \cap D) = 16/24 \neq P(M) \cdot P(D) = 18/24 \cdot 20/24 = 0.625 \longrightarrow$ abhängig

Zufallsvariable

Bildet Scharnier zwischen Statistik und Wahrscheinlichkeitsrechnung

X = Zufallsvariable mit Werten $x_1, x_2, \ldots$ und Wahrscheinlichkeit $P(X=x_1), P(X=x_2), \ldots$

z.B. Gefäss mit 4 von 1-4 nummerierten Kugeln. Nacheinander 2 ziehen ohne zurücklegen

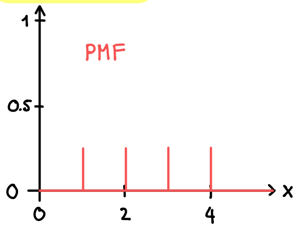
X = Nummer 2. gezogener Kugel

x	1	2	3	4
$f_X(x) = P(X=x)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

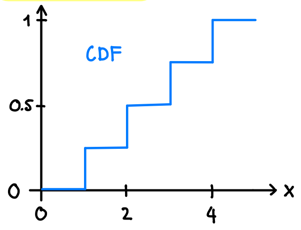
Z = Summe beider gezogenen Kugel

z	3	4	5	6	7
$f_Z(z) = P(Z=z)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{6}$

$f_X(x) = P(X=x)$



$F_X(x) = P(X \leq x)$



Standardisierte Zufallsvariable

Mit Zufallsvariable X, Erwartungswert $E(X)$, Standardabweichung $S(X)$

$W = \frac{X - E(X)}{S(X)} = \frac{1}{S(X)} \cdot X - \frac{E(X)}{S(X)}$ ⓘ $E(W) = 0$ und $S(W) = 1$

z.B. Vorheriges Beispiel mit Kugeln und Zufallsvariable X

x	1	2	3	4
$P(X=x)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

→

$w = \frac{X - 2.5}{1.118}$	-1.34	-0.45	0.45	1.34
$P(W=w)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

↳ $E(W) = 0$ und $S(W) = 1$

Linearkombination von Zufallsvariablen

$E(a \cdot X + b \cdot Y + c \cdot Z + \ldots) = a \cdot E(X) + b \cdot E(Y) + c \cdot E(Z)$

$V(a \cdot X + b \cdot Y + c \cdot Z + \ldots) = a^2 \cdot V(X) + b^2 \cdot V(Y) + c^2 \cdot V(Z)$ ⓘ Nur wenn paarweise unabhängig

z.B. X, Y, Z paarweise unabhängig mit $E(X) = 5$, $E(Y) = 10$, $E(Z) = 15$

und $V(X) = 1$, $V(Y) = 4$, $V(Z) = 2$

$E(X + \frac{1}{2} Y - 2 \cdot Z) = E(X) + \frac{1}{2} \cdot E(Y) - 2 \cdot E(Z) = 5 + \frac{1}{2} \cdot 10 - 2 \cdot 15 = -20$

$V(X + \frac{1}{2} Y - 2 \cdot Z) = V(X) + (\frac{1}{2})^2 \cdot V(Y) + (-2)^2 \cdot V(Z) = 1 + \frac{1}{4} \cdot 4 + 4 \cdot 2 = 10$

Kennzahlen einer Zufallsvariable

Erwartungswert (Mittelwert): $E(X) = P(X=x_1) \cdot x_1 + P(X=x_2) \cdot x_2 + \ldots$

Varianz: $V(X) = E(X^2) - (E(X))^2$

Standardabweichung: $S(X) = \sqrt{V(X)}$

z.B. Vorheriges Beispiel mit Kugeln und Zufallsvariable X

$E(X) = \frac{1}{4} \cdot 1 + \frac{1}{4} \cdot 2 + \frac{1}{4} \cdot 3 + \frac{1}{4} \cdot 4 = 2.5$

$E(X^2) = \frac{1}{4} \cdot 1^2 + \frac{1}{4} \cdot 2^2 + \frac{1}{4} \cdot 3^2 + \frac{1}{4} \cdot 4^2 = 7.5$

$V(X) = 7.5 - 2.5^2 = 1.25$

$S(X) = \sqrt{1.25} \approx 1.118$

Lineare Transformation einer Zufallsvariable

Translation einer Zufallsvariable X um die Zahl b:

$E(X+b) = E(X) + b$

$V(X+b) = V(X)$

$S(X+b) = S(X)$

Skalierung einer Zufallsvariable X mit die Zahl a:

$E(a \cdot X) = a \cdot E(X)$

$V(a \cdot X) = a^2 \cdot V(X)$

$S(a \cdot X) = |a| \cdot S(X)$

z.B. Vorheriges Beispiel mit Kugeln und Zufallsvariable X

$$Y = X + 4 \quad \begin{cases} E(Y) = 2.5 + 4 = 7.5 \\ V(Y) = 1.25 \\ S(Y) = 1.118 \end{cases}$$

$$Y = 2 \cdot X \quad \begin{cases} E(Y) = 2 \cdot 2.5 = 5 \\ V(Y) = 2^2 \cdot 1.25 = 5 \\ S(Y) = |2| \cdot 1.118 = 2.236 \end{cases}$$

2 Zufallsvariablen

$E(X+Y) = E(X) + E(Y)$

$E(X \cdot Y) = E(X) \cdot E(Y)$

$V(X+Y) = V(X) + V(Y)$

$V(X \cdot Y) = V(X) \cdot V(Y)$

ⓘ Nur wenn beide unabhängig

Faires Spiel

Wenn $E(\text{Auszahlung})$ = Einsatz

z.B. $P(\text{2-Mal ziehen}) = \frac{1}{6}$, 10.- Auszahlung

$P(\text{3-Mal ziehen}) = \frac{1}{3}$, Einsatz wird ausgezahlt

$P(\text{4-Mal ziehen}) = \frac{1}{2}$, 0.- Auszahlung

Ist Spiel fair mit Einsatz von 4.- ?

↳ $E(\text{Auszahlung}) = \frac{1}{6} \cdot 10 + \frac{1}{3} \cdot 4 + \frac{1}{2} \cdot 0 = 3 \neq 4$

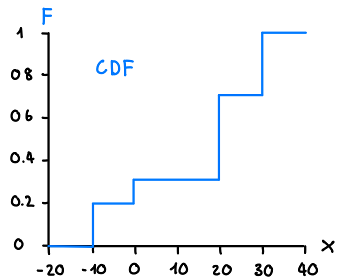
→ Spiel ist nicht fair

Für welchen Einsatz ist Spiel fair?

↳ $E(X) = \frac{1}{6} \cdot 10 + \frac{1}{3} \cdot X + \frac{1}{2} \cdot 0 = X$

$\frac{5}{3} = \frac{2}{3} \cdot X \rightarrow 2.5 = X$

CDF und Wahrscheinlichkeit



$P(X \leq 10) = 0.3$, $P(X \leq 20) = 0.7$

$P(X = 10) = 0$, $P(X = 20) = 0.4$

$P(X < 10) = 0.3$, $P(X < 20) = 0.3$

$P(X > 10) = 0.7$, $P(X > 20) = 0.3$

$P(10 < X \leq 20) = 0.4$, $P(10 \leq X < 20) = 0$

Wahrscheinlichkeitsverteilungen

Bernoulli-Verteilung

Bernoulli-Zufallsvariable wenn X nur Wert 0 oder 1 annimmt

$\left.\begin{array}{l}\text{Erfolgswahrscheinlichkeit } p \text{ für Wert } 1\\ \text{Misserfolgswahrscheinlichkeit } q \text{ für Wert } 0\end{array}\right\} q = 1-p$

$E(X) = p$ und $V(X) = p \cdot q$

z.B. Einmaliger Wurf von gewöhnlichem Würfel

$$X \begin{cases} 1 \text{ wenn Augenzahl gleich } 6 \\ 0 \text{ wenn Augenzahl ungleich } 6 \end{cases}$$

$p = 1/6,\ q = 5/6,\ E(X) = 1/6,\ V(X) = 1/6 \cdot 5/6 = 5/36$

Binomialverteilung

Wenn $X = X_1 + \ldots + X_n$ wobei $X_1 + \ldots + X_n$ n voneinander unabhängige Bernoulli-Zufallsvariablen mit gleicher Erfolgswahrscheinlichkeit p sind

$X \sim \text{Binom}(n, p)$

$P(X=x) = \binom{n}{x} \cdot p^x \cdot q^{n-x}$ ⓘ PMF

$E(X) = n \cdot p$, $V(X) = n \cdot p \cdot q$ und $S(X) = \sqrt{V(X)}$

z.B. Multiple-Choice Test mit 9 Fragen. Pro Frage 3 Antwortmöglichkeiten wovon 1 richtig.
Alle Fragen unabhängig beantwortbar. Test bei mind. 6 richtigen Antworten bestanden.

$$X_i \begin{cases} 1 \text{ wenn i-te Frage richtig} \\ 0 \text{ wenn i-te Frage falsch} \end{cases} \quad i = 1, \ldots, 9$$

$X = X_1 + \ldots + X_9$ = Anzahl Fragen richtig

$X \sim \text{Binom}(9, 1/3)$

$E(X) = 9 \cdot 1/3 = 3$, $V(X) = 9 \cdot 1/3 \cdot 2/3 = 2$ und $S(X) = \sqrt{2} \approx 1.4$

$P(X=x) = \binom{9}{x} \cdot (1/3)^x \cdot (2/3)^{9-x}$

X	0	1	2	3	4	5	6	7	8	9
$P(X=x)$	.026	.117	.234	.273	205	.102	034	.007	.000	.000

Wahrscheinlichkeit für Bestehen: $P(X \geq 6) = P(X=6) + \ldots + P(X=9) = 0.041$

z.B. Würfel mit 1-Mal 1, 2-Mal 2 und 3-Mal 3, 12-Mal werfen

Augenzahl 1: $X \sim \text{Binom}(12, 1/6)$

→ genau 2-Mal: $P(X=2) = \binom{12}{2} \cdot (1/6)^2 \cdot (5/6)^{10} = \underline{\underline{0.29}}$

→ max. 2-Mal: $P(X \leq 2) = P(X=0) + P(X=1) + P(X=2)$
$\approx 0.1122 + 0.2692 + 0.2961 \approx \underline{\underline{0.67}}$

Augenzahl 2: $X \sim \text{Binom}(12, 1/3)$

→ mind. 3 und max. 5: $P(3 \leq X \leq 5) = P(X=3) + P(X=4) + P(X=5)$
$\approx 0.212 + 0.2384 + 0.1908 \approx \underline{\underline{0.64}}$

Augenzahl 3: $X \sim \text{Binom}(12, 1/2)$

→ mind. 3-Mal: $P(X \geq 3) = 1 - P(X \leq 2) = 1 - (P(X=0) - P(X=1) - P(X=2))$
$\approx 1 - (0.0002 - 0.0029 - 0.0161) \approx \underline{\underline{0.98}}$

z.B. Wie oft muss gewöhnlicher Würfel geworfen werden, damit Wahrscheinlichkeit,

dass immer gerade Augenzahl auftritt 0.5 % beträgt

→ $X \sim \text{Binom}(n, 1/2)$, $P(X=n) = \underbrace{\binom{n}{n}}_{1} \cdot (1/2)^n \cdot \underbrace{(1/2)^{n-n}}_{1} = 0.005 \longrightarrow n = \log_{1/2}(0.005) \approx 7.64 \longrightarrow$ 8-Mal

TR $\log(0.005) : \log(0.5)$

dass mind. 1-Mal Augenzahl 6 auftritt 99 % beträgt

→ $X \sim \text{Binom}(n, 1/6)$, $P(X \geq 1) = 1 - P(X=0) = 1 - \underbrace{\binom{n}{0}}_{1} \cdot \underbrace{(1/6)^0}_{1} \cdot (5/6)^{n-0} = 0.99 \longrightarrow n = \log_{5/6}(0.01) \approx 25.26 \longrightarrow$ 26-Mal

TR $\log(0.01) : \log(5/6)$

Poisson-Verteilung

Die zu zählenden Ereignisse treten mit einer konstanten mittleren Rate $\lambda > 0$ ein.

$X \sim \text{Poisson}(\lambda)$

$P(X=x) = e^{-\lambda} \cdot \frac{\lambda^x}{x!}$ ⓘ PMF, e = Eulersche Zahl

$E(X) = \lambda$, $V(X) = \lambda$ und $S(X) = \sqrt{V(X)}$

z.B. Poisson-verteilte Zufallsvariable X mit Parameter $\lambda = 4$

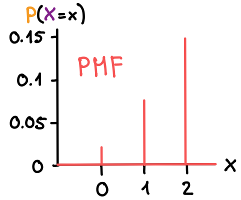
$X \sim \text{Poisson}(4)$

$P(X=0) = e^{-4} \cdot \frac{4^0}{0!} \approx 0.015$

$P(X=1) = e^{-4} \cdot \frac{4^1}{1!} \approx 0.073$

$P(X=2) = e^{-4} \cdot \frac{4^2}{2!} \approx 0.147$

usw.



$E(X) = 4$, $V(X) = 4$, $S(X) = \sqrt{4} = 2$

Approximation von Binomialverteilung durch Poissonverteilung

Binomialverteilte Zufallsvariable $X \sim \text{Binom}(n, p)$ kann für $n \geq 50$ und $p \leq 0.1$ durch Poissonverteilte Zufallsvariable $Y \sim \text{Poisson}(\lambda)$ mit $\lambda = n \cdot p$ approximiert werden.

z.B. Auf Flug mit 200 gebuchten Plätzen erscheinen 5 % nicht. Wahrscheinlichkeit für genau 8 No-Shows:

Exakt: $X \sim \text{Binom}(200, 0.05) \rightarrow P(X=8) = \binom{200}{8} \cdot 0.05^8 \cdot 0.95^{200-8\ (=192)} = 0.1137$

Approximativ: $Y \sim \text{Poisson}(200 \cdot 0.05\ (=10)) \longrightarrow P(Y=8) = e^{-10} \cdot \frac{10^8}{8!} = 0.1126$

Stetige Zufallsvariablen

Wenn X jede reelle Zahl aus Intervall als Wert annimmt

① Nicht stetige Zufallsvariablen heissen diskret

$P(X=x)=0$ für alle x Werte

$P(X \le x) = P(X<x)$ / $P(X \ge x) = P(X>x)$

$P(a \le X \le b) = P(a \le X < b) = P(a < X \le b) = P(a<X<b)$ für alle $a<b$

Erwartungswert μ: $E(X) = \int_{-\infty}^{\infty} f(x) \cdot x \, dx$

Varianz σ^2: $V(X) = E(X^2) - E(X)^2$ mit $\int_{-\infty}^{\infty} f(x) \cdot x^2 \, dx$

Standardabweichung σ: $S(X) = \sqrt{V(X)}$

$E(X+Y) = E(X) + E(Y)$, $E(\alpha \cdot X) = \alpha \cdot E(X)$

$V(\alpha \cdot X + \beta) = \alpha^2 \cdot V(X)$

$V(X+Y) = V(X) + V(Y)$ für unabhängige X, Y

Intervallwahrscheinlichkeiten

$P(X \le x) = F(x) = \int_{-\infty}^{x} f(u) \, du$ ① Vorzeichen von x übernehmen

$P(a < X \le b) = P(X \le b) - P(X \le a) = F(b) - F(a) = \int_a^b f(u) \, du$

$P(X > x) = 1 - P(X \le x) = 1 - F(x) = \int_x^{\infty} f(u) \, du$

z.B. $F(x) = -0.25x^3 + 0.75x + 0.5$

$P(-0.5 \le X \le 0.5) = F(0.5) - F(-0.5) = 0.84375 - 0.15625 = 0.6875$

$P(X > 0.8) = 1 - F(0.8) = 1 - 0.972 = 0.028$

Dichtefunktion PDF f(x)

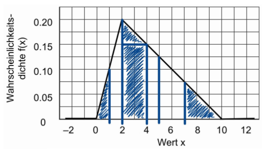
Wenn X eine stetige Zufallsvariable mit CDF F(x)

PDF $\rightarrow$ $F(x) = P(X \le x) = \int_{-\infty}^{x} f(u) \, du$

$f(x)$ ist Ableitung von $F(x)$: $f(x) = F'(x)$

Es gilt immer: $f(x) \ge 0$ und $\int_{-\infty}^{\infty} f(x) \, dx = 1$

z.B.



Fläche Dreieck mit Grundlinie 10 und Höhe 0.2

$\int_{-\infty}^{\infty} f(x) \, dx = \frac{1}{2} \cdot 10 \cdot 0.2 = 1$

$P(X \le 1) = \frac{1}{2} \cdot 1 \cdot 0.1 = 0.05$

$P(2 \le X \le 4) = 2 \cdot 0.15 + \frac{1}{2} \cdot 2 \cdot 0.05 = 0.35$

$P(X=5) = 0$

$P(X \ge 7) = \frac{1}{2} \cdot 3 \cdot 0.075 = 0.1125$

Kumulative Verteilungsfunktion CDF F(x)

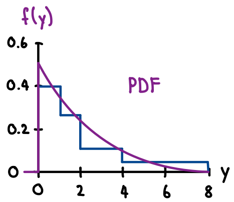
Wenn X eine stetige Zufallsvariable mit Wahrscheinlichkeitsmass P

CDF $\rightarrow$ $F(x) = P(X \le x)$

Beispiel

Brenndauer in Jahren	[0,1[	[1,2[	[2,4[	[5,8]
Klassenbreite b_k	1	1	2	3
relative Häufigkeit P_k	0.4	0.25	0.25	0.1
relative Häufigkeitsdichte $\frac{P_k}{b_k}$	0.4	0.25	0.125	0.025

$$f(y) \begin{cases} 0 & \text{für } y<0 \\ \frac{1}{2} \cdot e^{-\frac{1}{2} \cdot y} & \text{für } y \ge 0 \end{cases}$$



Überprüfungen:

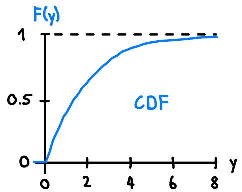
$f(y) \ge 0$ für alle y ✓

$$\int_{-\infty}^{\infty} f(y) \, dy = \int_0^{\infty} \frac{1}{2} \cdot e^{-\frac{1}{2} \cdot y} \, dy = \left[\frac{1/2}{-1/2} \cdot e^{-\frac{1}{2} \cdot y}\right]_0^{\infty} = (-e^{-\frac{1}{2} \cdot \infty}) - (-e^{-\frac{1}{2} \cdot 0}) = 1$$ ✓

$P(Y \le y) = F(y) = \int_{-\infty}^{y} f(u) \, du$

$= \int_0^y \frac{1}{2} \cdot e^{-\frac{1}{2} \cdot u} \, du = \left[-e^{-\frac{1}{2} \cdot y}\right]_0^y$

$= (-e^{-\frac{1}{2} \cdot y}) - (-e^{-\frac{1}{2} \cdot 0}) = 1 - e^{-\frac{1}{2} \cdot y}$



$$F(y) \begin{cases} 0 & \text{für } y<0 \\ 1 - e^{-\frac{1}{2} \cdot y} & \text{für } y \ge 0 \end{cases}$$

Partielle Integration: $\int_a^b u(x) \cdot v'(x) \, dx = \left[u(x) \cdot v(x)\right]_a^b - \int_a^b u'(x) \cdot v(x) \, dx$

$E(Y) = \int_{-\infty}^{\infty} f(x) \cdot y \, dx = \int_0^{\infty} \frac{1}{2} \cdot e^{-\frac{1}{2} \cdot y} \cdot y \, dy$

$= \left[-e^{-\frac{1}{2} \cdot y} \cdot y\right]_0^{\infty} - \int_0^{\infty} -e^{-\frac{1}{2} \cdot y} \cdot 1 \, dy = \left[-e^{-\frac{1}{2} \cdot y} \cdot y\right]_0^{\infty} - \left[2 \cdot e^{-\frac{1}{2} \cdot y}\right]_0^{\infty}$

$= \left((-e^{-\frac{1}{2} \cdot \infty} \cdot \infty) - (-e^{-\frac{1}{2} \cdot 0} \cdot 0)\right) - \left((2 \cdot e^{-\frac{1}{2} \cdot \infty}) - (2 \cdot e^{-\frac{1}{2} \cdot 0})\right) = \underline{\underline{2}}$

$V(Y) = E(Y^2) - E(Y)^2 = \int_0^{\infty} \frac{1}{2} \cdot e^{-\frac{1}{2} \cdot y} \cdot y^2 \, dy - 2^2 = \dots = \underline{\underline{4}}$

$S(Y) = \sqrt{V(Y)} = \sqrt{4} = \underline{\underline{2}}$

Integrieren / Ableiten

x^n	$\frac{1}{n+1} \cdot x^{n+1}$		x^n	$n \cdot x^{n-1}$
e^x	e^x		e^x	e^x
e^{-x}	$-e^{-x}$		e^{-x}	$-e^{-x}$
e^{-cx}	$-\frac{1}{c} \cdot e^{-cx}$		e^{2x}	$2 \cdot e^{2x}$
			e^{x^2}	$2x \cdot e^{x^2}$

Integrieren: $f'(x) \rightarrow f(x)$, $f(x) \rightarrow F(x)$

Ableiten: $F(x) \rightarrow f(x)$, $f(x) \rightarrow f'(x)$

Normalverteilung / Gauss-Verteilung

Wenn stetige Zufallsvariable X mit μ und σ folgende PDF besitzt:

$$f(x) = \frac{1}{\sqrt{2 \cdot \pi} \; \sigma} \cdot e^{-\frac{1}{2} \cdot \left(\frac{x-\mu}{\sigma}\right)^2}$$

Schreibweise: $X \sim \text{Normal}(\mu, \sigma)$ oder kurz $X \sim N(\mu, \sigma)$

$E(X) = \mu$, $V(X) = \sigma^2$, $S(X) = \sigma$

$P(\mu - \sigma \le X \le \mu + \sigma) \approx 0.68$

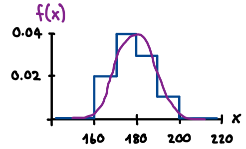
$P(\mu - 2 \cdot \sigma \le X \le \mu + 2 \cdot \sigma) \approx 0.95$

$P(\mu - 3 \cdot \sigma \le X \le \mu + 3 \cdot \sigma) \approx 0.997$

z.B. 10 Körpergrössen mit $\mu = 178$ und $\sigma = 10$

$X \sim N(178, 10)$ mit

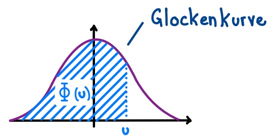
$$f(x) = \frac{1}{\sqrt{2 \cdot \pi} \; 10} \cdot e^{-\frac{1}{2} \cdot \left(\frac{x-178}{10}\right)^2}$$



CDF Normalverteilung

CDF $\rightarrow$ $\Phi(u) = \frac{1}{\sqrt{2 \cdot \pi}} \cdot \int_{-\infty}^{u} e^{-\frac{1}{2} \cdot \left(\frac{x-\mu}{\sigma}\right)^2} dx$

① Werte von $\Phi(u)$ aus Tabelle entnehmen



Linearkombination Normalverteilung

Wenn $X \sim N(\mu_X, \sigma_X)$, $Y \sim N(\mu_Y, \sigma_Y)$, α und β Zahlen (nicht beide 0)

Dann gilt für Zufallsvariable $Z = \alpha \cdot X + \beta \cdot Y$.

Z ist immer auch normalverteilt, also $Z \sim N(\mu_Z, \sigma_Z)$

$\mu_Z = \alpha \cdot \mu_X + \beta \cdot \mu_Y$, $\sigma_Z = \sqrt{\alpha^2 \cdot \sigma_X^2 + \beta^2 \cdot \sigma_Y^2}$ (σ_Z nur wenn X und Y unabhängig)

z.B. unabhängig: $X \sim N(0,1)$, $Y \sim N(1,2)$ und $\alpha = 3$, $\beta = 2$

$Z = 3 \cdot X + 2 \cdot Y \sim N(3 \cdot 0 + 2 \cdot 1, \sqrt{3^2 \cdot 1^2 + 2^2 \cdot 2^2}) \sim N(2, 5)$

Standardnormalverteilung

Wenn X normalverteilte Zufallsvariable mit $\mu = 0$ und $\sigma = 1$, also $X \sim N(0,1)$

Dann lautet PDF von X: $\varphi(x) = \frac{1}{\sqrt{2 \cdot \pi}} \cdot e^{-\frac{1}{2} \cdot x^2}$

Standardnormalverteilungstabelle

U muss standardnormalverteilte Zufallsvariable sein, $U \sim N(0,1)$

Ansonsten $X \sim N(\mu, \sigma) \rightarrow U = \frac{X-\mu}{\sigma} \sim N(0,1)$

$P(U \leq u) = \Phi(u)$ / $P(U \leq -u) = 1 - \Phi(u)$ ⓘ Vorzeichen von u nicht übernehmen

$P(U \geq u) = 1 - \Phi(u)$ / $P(U \geq -u) = \Phi(u)$

$P(-u \leq U \leq u) = 2 \cdot \Phi(u) - 1$ / $P(-a \leq U \leq b) = \Phi(b) + \Phi(a) - 1$ / $P(a \leq U \leq b) = \Phi(b) - \Phi(a)$

$\Phi(-u) = 1 - \Phi(u)$

z.B. $P(U \leq -1)$ mit $U \sim N(0,1)$

$= 1 - \Phi(1) = 1 - 0.8413 = \underline{\underline{0.1587}}$

$P(X \leq 5)$ mit $X \sim N(3,1)$

X nicht standardnormalverteilt $\rightarrow U = \frac{X-3}{1} = X-3$

$= P(X \leq 5-3) = P(U \leq 2) = \Phi(2) = \underline{\underline{0.9772}}$

$P(-0.5 \leq U \leq 0.5)$ mit $U \sim N(0,1)$

$= 2 \cdot \Phi(0.5) - 1 = 2 \cdot 0.6915 - 1 = \underline{\underline{0.383}}$

Stetigkeitskorrektur

Wenn diskrete Verteilung durch Normalverteilung approximiert wird

ⓘ Nicht nötig bei Intervall oder Durchschnitt ($\overline{Z_n}$)

$P(Z \leq b) \approx P(Z \leq b + 0.5)$

$P(Z \geq a) \approx P(Z \geq a - 0.5)$

$P(a \leq Z \leq b) \approx P(a - 0.5 \leq Z \leq b + 0.5)$

$\leq \neq <$ und $\geq \neq >$ bei diskretem Z

z.B. $P(X < 58)$ zu $P(X \leq 57 + 0.5)$

$P(X > 58)$ zu $P(X \geq 59 - 0.5)$

Approximation Binomialverteilung durch Normalverteilung

Wenn Zufallsvariablen X binomialverteilt mit n und p, also $X \sim B(n,p)$

Dann gilt für normalverteilte Zufallsvariable $Z \sim N(n \cdot \mu, \sqrt{n} \cdot \sigma)$:

μ: p, σ: $\sqrt{p \cdot (1-p)}$

z.B. n-Mal Würfel mit Augenzahlen 1,2,3 (auf 2 Seiten) werfen

wobei $Z_i \begin{cases} 1 & \text{wenn Augenzahl 3} \\ 0 & \text{sonst} \end{cases}$, $Z_1, \ldots, Z_n$ Bernoulli-verteilt mit $p = 1/3$

$\mu = p = 1/3$, $\sigma = \sqrt{p \cdot (1-p)} = \sqrt{2}/3$

Appoximative Berechnung von $P(Z_{24} \leq 10)$

$Z_{24} = X_1 + \ldots + X_n$: $E(Z_{24}) = n \cdot \mu = 24 \cdot 1/3 = 8$, $S(Z_{24}) = \sqrt{n} \cdot \sigma = \sqrt{24} \cdot \sqrt{2}/3 \approx 2.309$

$Z_{24} \sim N(8, 2.309)$ nicht standardnormalverteilt $\rightarrow U = \frac{Z_{24} - 8}{2.309}$

$= P(Z_{24} \leq 10 + 0.5) = P(Z_{24} \leq \frac{10.5 - 8}{2.309}) = P(Z_{24} \leq 1.08) = \Phi(1.08) = \underline{\underline{0.8599}}$

Zentraler Grenzwertsatz

Mittelwert und Standardabweichung sind für grosse n näherungsweise normalverteilt unabhängig der ursprünglichen Verteilung

Wenn Zufallsvariablen $X_1, \ldots, X_n$ identisch verteilt und paarweise unabhängig mit $\mu = E(X_1) = \ldots = E(X_n)$ und $\sigma = S(X_1) = \ldots = S(X_n)$

Dann gilt für Zufallsvariable $Z_n = X_1 + \ldots + X_n$:

$E(Z_n) = n \cdot \mu$, $S(Z_n) = \sqrt{n} \cdot \sigma$, $V(Z_n) = n \cdot \sigma^2$

$Z_n \sim N(n \cdot \mu, \sqrt{n} \cdot \sigma)$

Dann gilt für Zufallsvariable $\overline{Z_n} = \frac{X_1 + \ldots + X_n}{n}$

$E(\overline{Z_n}) = \mu$, $S(\overline{Z_n}) = \frac{\sigma}{\sqrt{n}}$, $V(\overline{Z_n}) = \frac{\sigma^2}{n}$

$Z_n \sim N(\mu, \frac{\sigma}{\sqrt{n}})$

z.B. n-Mal zufällige reelle Zahl aus Intervall [-1, 1] wählen

wobei Z_i = i-te gewählte Zahl, $Z_1, \ldots, Z_n$ stetig gleichverteilt

$\mu = \frac{-1+1}{2} = 0$, $\sigma = \sqrt{\frac{(1-(-1))^2}{12}} = \frac{1}{\sqrt{3}}$

Appoximative Berechnung von $P(-0.1 \leq \overline{Z_{48}} \leq 0.1)$

$\overline{Z_{48}} = \frac{X_1 + \ldots + X_{48}}{48}$: $E(\overline{Z_{48}}) = \mu = 0$, $S(\overline{Z_{48}}) = \frac{\sigma}{\sqrt{n}} = \frac{1/\sqrt{3}}{\sqrt{48}} = \frac{1}{12}$

$\overline{Z_{48}} \sim N(0, \frac{1}{12})$ nicht standardnormalverteilt $\rightarrow U = \frac{\overline{Z_{48}} - 0}{1/12} = \frac{\overline{Z_{48}}}{1/12}$

$= P\left(\frac{-0.1}{1/12} \leq \overline{Z_{48}} \leq \frac{0.1}{1/12}\right) = P(-1.2 \leq U \leq 1.2) = 2 \cdot \Phi(1.2) - 1 = 2 \cdot 0.8849 - 1 = \underline{\underline{0.7698}}$

Berechnung von n, so dass $P(-0.1 \leq \overline{Z_n} \leq 0.1) = 0.9$

$E(\overline{Z_n}) = \mu = 0$, $S(\overline{Z_n}) = \frac{\sigma}{\sqrt{n}} = \frac{1/\sqrt{3}}{\sqrt{n}} = \frac{1}{\sqrt{3 \cdot n}}$

$\overline{Z_n} \sim N(0, 1/\sqrt{3 \cdot n})$ nicht standardnormalverteilt $\rightarrow U = \frac{\overline{Z_n} - 0}{1/\sqrt{3 \cdot n}} = \frac{\overline{Z_n}}{1/\sqrt{3 \cdot n}}$

$P(-0.1 \leq \overline{Z_n} \leq 0.1) = 0.9 \rightarrow 2 \cdot \Phi(u) - 1 = 0.9 \rightarrow \Phi(u) = 0.95$

$\xrightarrow{\text{Tabelle}} u = \frac{1.64 + 1.65}{2} = 1.645 \rightarrow \frac{0.1}{1/\sqrt{3 \cdot n}} = 1.645 \rightarrow n = \frac{(1.645 : 0.1)^2}{3} \approx \underline{\underline{90}}$

z.B. Berechnung von μ, so dass $P(Z < 500) = 0.02$ ⓘ 0.02 nicht in Tabelle, daher $P(Z = -u)$

$E(Z) = \mu$, $S(Z) = 1$,

$Z \sim N(\mu, 1)$ nicht standardnormalverteilt $\rightarrow U = \frac{Z - \mu}{1} = Z - \mu$

$P(Z < 500) = 0.002 \rightarrow P(U \leq -u) = 1 - \Phi(u) = 0.002 \rightarrow \Phi(u) = 0.998$

$\xrightarrow{\text{Tabelle}} u = 2.88 \rightarrow 500 - \mu = 2.88 \rightarrow \mu = \underline{\underline{502.88}}$

z.B. Berechnung von σ, so dass $P(-3 \leq Z \leq 3) = 0.99$

$E(Z) = 0$, $S(Z) = \sigma$

$Z \sim N(0, \sigma)$ nicht standardnormalverteilt $\rightarrow U = \frac{Z - 0}{\sigma} = \frac{Z}{\sigma}$

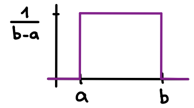
$P(-3 \leq Z \leq 3) = 0.99 \rightarrow 2 \cdot \Phi(u) - 1 = 0.99 \rightarrow \Phi(u) = 0.995$

$\xrightarrow{\text{Tabelle}} u = \frac{2.57 + 2.58}{2} = 2.575 \rightarrow \frac{3}{\sigma} = 2.575 \rightarrow \sigma = \underline{\underline{1.165}}$

Gleichverteilung

Wenn stetige Zufallsvariable X folgende PDF besitzt mit $a < b$:

$f(x) \begin{cases} \frac{1}{b-a} & \text{für } a \leq x \leq b \\ 0 & \text{für } x < a \text{ und } x > b \end{cases}$



Schreibweise: $X \sim \text{Uniform}(a, b)$

(gleichverteilt auf $[a, b]$)

$E(X) = \frac{a+b}{2}$, $V(X) = \frac{(b-a)^2}{12}$

z.B. $X \sim \text{Uniform}(0, 2\pi)$

$E(X) = \frac{0 + 2\pi}{2} = \pi$

$V(X) = \frac{(2 \cdot \pi - 0)^2}{12} = \frac{\pi^2}{3}$

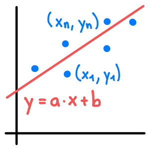
Regression

Lineare Regression

Gegeben: Datenpaare $(x_1, y_1), \ldots, (x_n, y_n)$ mit linearen Zusammenhang

Gesucht: Gerade $y = a \cdot x \cdot b$, welche Datenpaare möglichst gut approximiert

Bestmögliche Gerade heisst Regressionsgerade



Methode der kleinsten Quadrate

Gegeben: Datenpaare und eine Gerade $y = a \cdot x \cdot b$

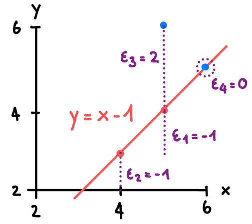
Dann heissen für $i = 1, 2, \ldots, n$:

- y_i beobachtete/totale y-Werte
- $\hat{y}_i = a \cdot x_i + b$ prognostizierte/erklärte y-Werte
- $\varepsilon_i = y_i - \hat{y}_i$ Residuen

Gerade, bei der Summe der Quadrate aller Residien $\varepsilon_1 + \varepsilon_2 + \ldots + \varepsilon_n$ am kleinsten ist, heisst Regressionsgerade

z.B. Residien bei Gerade $y = x - 1$

i	x_i	y_i	$\hat{y}_i = x_i - 1$	$\varepsilon_i = y_i - \hat{y}_i$	ε_i^2
1	4	2	3	-1	1
2	5	3	4	-1	1
3	5	6	4	2	4
4	6	5	5	0	0
				Summe	6



Regressionsgerade

Verwendung: Prognose/Erklärung von Merkmal Y durch Merkmal X

Gegeben: Datenpaare mit Mittelwerte $\bar{x}, \bar{y}$, Varianzen $\tilde{v}_x, \tilde{v}_y$, Kovarianz $\tilde{v}_{xy}$

Dann gilt für Regressionsgerade $y = a \cdot x + b$:

Steigung $a = \frac{\tilde{v}_{xy}}{\tilde{v}_x}$, y-Achsenabschnitt $b = \bar{y} - a \cdot \bar{x}$

Summe der Quadrate der Residien: $\varepsilon_1^2 + \ldots + \varepsilon_n^2 = n \cdot \left(\tilde{v}_y - \frac{\tilde{v}_{xy}^2}{\tilde{v}_x}\right)$

z.B.

i	1	2	3	4	Summe	Mittelwert	
x_i	4	5	5	6	20	5	$\bar{x}$
y_i	2	3	6	5	16	4	$\bar{y}$
x_i^2	16	25	25	36	102	25.5	$\tilde{v}_x = 25.5 - 5^2 = 0.5$
y_i^2	4	9	36	25	74	18.5	$\tilde{v}_y = 18.5 - 4^2 = 2.5$
$x_i \cdot y_i$	8	15	30	30	83	20.75	$\tilde{v}_{xy} = 20.75 - (5 \cdot 4) = 0.75$

Standardabweichungen: $\tilde{s}_x = \sqrt{0.5} \approx 0.71$, $\tilde{s}_y = \sqrt{2.5} \approx 1.58$

Korrelationskoeffizient nach Pearson: $r_{xy} = \frac{\tilde{v}_{xy}}{\tilde{s}_x \cdot \tilde{s}_y} = \frac{0.75}{0.71 \cdot 1.58} \approx 0.67$

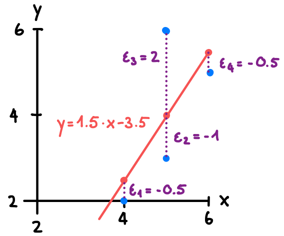
(!) $0.5 \leq r_{xy} \leq 0.75 \to$ x und y moderat positiv korreliert

Regressionsgerade $y = a \cdot x + b$ mit:

$a = \frac{\tilde{v}_{xy}}{\tilde{v}_x} = \frac{0.75}{0.5} = 1.5$

$b = \bar{y} - a \cdot \bar{x} = 4 - 1.5 \cdot 5 = -3.5$

$\varepsilon_1^2 + \ldots + \varepsilon_n^2 = n \cdot \left(\tilde{v}_y - \frac{\tilde{v}_{xy}^2}{\tilde{v}_x}\right) = 4 \cdot \left(2.5 \cdot \frac{0.75^2}{0.5}\right) = 5.5$



Varianzzerlegung

Gegeben: Regressionsgerade zu Datenpaare mit Mittelwerte $\bar{x}, \bar{y}$,

Varianzen $\tilde{v}_x, \tilde{v}_y$, Kovarianz $\tilde{v}_{xy}$, $\hat{y}_i = a \cdot x_i + b$ mit Mittelwert $\bar{\hat{y}}$

und Varianz $\tilde{v}_{\hat{y}}$, $\varepsilon_i = y_i - \hat{y}_i$ mit Mittelwert $\bar{\varepsilon}$ und Varianz $\tilde{v}_{\varepsilon}$

Dann gilt: Mittelwerte $\bar{y} = \bar{\hat{y}}$ und $\bar{\varepsilon} = 0$

Varianzen $\tilde{v}_y = \tilde{v}_{\hat{y}} + \tilde{v}_{\varepsilon}$ und $\tilde{v}_{\hat{y}} = \frac{\tilde{v}_{xy}^2}{\tilde{v}_x}$

Bestimmtheitsmass R^2

Verhältnis erklärter Varianz $\tilde{v}_{\hat{y}}$ zu totaler Varianz $\tilde{v}_y$

$R^2 = \frac{\tilde{v}_{\hat{y}}}{\tilde{v}_y} = r_{xy}^2$, $0 \leq R^2 \leq 1$

(!) 1 wenn alle Datenpaare auf Regressionsgerade

z.B. $0.5 \to 50\%$

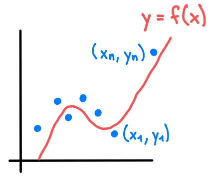
"R^2 der Gesamtvarianz in den y-Daten kann durch Regressionsgerade erklärt werden."

Nichtlineare Regression

Gegeben: Datenpaare $(x_1, y_1), \ldots, (x_n, y_n)$ mit nicht linearen Zusammenhang

Gesucht: Kurve $y = f(x)$, welche im Sinne der Methode der kleinsten Quadrate die Datenpaare möglichst gut annähert

(!) Mit Variablentransformation auf lineare Regression zurückführen



z.B. Exponentialfunktion: $y = a \cdot b^x$, Daten:

x_i	0	2	4	8
y_i	1	0.75	0.35	0.15

t_i	0	2	4	8
$z_i = \log_{10}(y_i)$	0	-0.12	-0.46	-0.82

Mittelwert: $\bar{t} = 3.5$, $\bar{z} = -0.35$

Varianz: $\tilde{v}_t = 8.75$, $\tilde{v}_z = 0.1021$

Kovarianz: $\tilde{v}_{tz} = -0.935$

Regressionsgerade $z = c \cdot t + d$ mit:

$c = \frac{\tilde{v}_{tz}}{\tilde{v}_t} = \frac{-0.935}{8.75} = -0.107$

$d = \bar{z} - c \cdot \bar{t} = -0.35 - (-0.107) \cdot 3.5 = 0.0245$

Also: $z = -0.107 \cdot t + 0.0245$

Regressionskurve $y = a \cdot b^x$ mit:

$a = 10^d = 10^{0.0245} \approx 1.06$

$b = 10^c = 10^{-0.107} \approx 0.78$

Also: $y \approx 1.06 \cdot 0.78^x$

Variablentransformation

Exponentialfunktion: $y = a \cdot b^x$

↳ $\overset{z}{\log_{10}(y)} = \overset{c}{\log_{10}(b)} \cdot \overset{t}{x} + \overset{d}{\log_{10}(a)}$

Potenzfunktion: $y = a \cdot x^m$

↳ $\overset{z}{\log_{10}(y)} = \overset{c}{m} \cdot \overset{t}{\log_{10}(x)} + \overset{d}{\log_{10}(a)}$

Quadratische Funktion: $y = a \cdot x^2 + b$

↳ $\overset{z}{y} = \overset{c}{a} \cdot \overset{t}{x^2} + \overset{d}{b}$ (mit $x \geq 0$)

Kehrwertfunktion: $y = \frac{1}{a \cdot x + b}$

↳ $\overset{z}{\frac{1}{y}} = \overset{c}{a} \cdot \overset{t}{x} + \overset{d}{b}$

Schliessende Statistik

Ziel: Von bekannter Stichprobe Rückschlüsse auf unbekannte Grundgesamtheit ziehen

z.B. Von bekanntem Stichprobenmittelwert $\bar{x}$ auf unbekannten Grundgesamtheitmittelwert μ schliessen

Punktschätzung

Mittelwert $\bar{x} = \frac{x_1 + \ldots + x_n}{n}$ der Stichprobe ist geeigneter Mittelwert μ der Grundgesamtheit.

Korrigierte Standardabweichung $s = \sqrt{\frac{x_1^2 + \ldots + x_n^2}{n} - \bar{x}^2}$ der Stichprobe ist geeignete Standardabweichung σ der Grundgesamtheit.

Intervallschätzung

Schätzintervall $[\Theta_u, \Theta_o]$ bilden, welches μ, σ^2, p mit vorgegebener Sicherheit γ enthält.

Begriffe: Schätzintervall = Vertrauensintervall = Konfidenzintervall

Sicherheit = Vertrauensniveau = Konfidenzlevel

(!) Je höher das Konfidenzlevel, desto breiter das Konfidenzintervall.

↳ Wollen mehr Sicherheit, dass Intervall wahren Wert enthält → breiteres Intervall

Je grösser der Stichprobenumfang, desto enger das Konfidenzintervall.

↳ Grösserer Stichprobenumfang gibt mehr Informationen und reduziert Unsicherheit → engerer Intervall

	(1) Verteilung der Grundgesamtheit	(2) Param.	(3) Schätzfunktionen	(4) zugehörige standardisierte Zufallsvariable	(5) Verteilung und benötigte Quantile	(6) Zufallsvariablen für Intervallgrenzen
1	Normalverteilung (Varianz σ^2 bekannt)	μ	$\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i$	$U = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$	**Standardnormalverteilung** (Tabelle 2) $c = u_p$ mit $p = \frac{1+\gamma}{2}$	$\Theta_u = \bar{X} - c \cdot \frac{\sigma}{\sqrt{n}}$ $\Theta_o = \bar{X} + c \cdot \frac{\sigma}{\sqrt{n}}$
2	Normalverteilung (Varianz σ^2 unbekannt und $n \le 30$; sonst Fall 1 mit s als Schätzwert für σ)	μ	$\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i$ $S^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (X_i - \bar{X})^2$	$T = \frac{\bar{X} - \mu}{S / \sqrt{n}}$	***t*-Verteilung** (Tabelle 4) mit $f = n - 1$ $c = t_{(p;f)}$ mit $p = \frac{1+\gamma}{2}$	$\Theta_u = \bar{X} - c \cdot \frac{S}{\sqrt{n}}$ $\Theta_o = \bar{X} + c \cdot \frac{S}{\sqrt{n}}$
3	Normalverteilung	σ^2	$\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i$ $S^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (X_i - \bar{X})^2$	$Z = (n-1) \frac{S^2}{\sigma^2}$	**Chi-Quadrat-Verteilung** (Tabelle 3) mit $f = n - 1$ $c_1 = z_{(p_1;f)}$ mit $p_1 = \frac{1-\gamma}{2}$ $c_2 = z_{(p_2;f)}$ mit $p_2 = \frac{1+\gamma}{2}$	$\Theta_u = \frac{(n-1) \cdot S^2}{c_2}$ $\Theta_o = \frac{(n-1) \cdot S^2}{c_1}$
4	Bernoulli-Verteilung Anteilsschätzung (mit $n\hat{p}(1-\hat{p}) > 9$)	p	$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ X_i 0/1-wertig mit $P(X_i = 1) = p$	$U = \frac{\bar{X} - p}{\sqrt{p(1-p)/n}}$	**Standardnormalverteilung** (näherungsweise), Tabelle 2 $c = u_q$ mit $q = \frac{1+\gamma}{2}$	$\Theta_u = \bar{X} - c \cdot \sqrt{\frac{\bar{X} \cdot (1 - \bar{X})}{n}}$ $\Theta_o = \bar{X} + c \cdot \sqrt{\frac{\bar{X} \cdot (1 - \bar{X})}{n}}$
5	beliebig mit $n > 30$	μ, σ^2	wie im Fall 1 (gegebenenfalls mit s als Schätzwert für σ) bzw. im Fall 3			

z.B. Gegeben: 27 defekte Teile in Stichprobe von Umfang $n = 500$

Gesucht: Schätzintervall für p mit Vertrauensniveau von 95 %

p gesucht → Fall 4

1. $\hat{p} = \bar{x} = \frac{27}{500} = 0.054$ → Faustregel prüfen: $500 \cdot 0.054(1 - 0.054) = 25.542 > 9$ ✓

2. $p = \frac{1+0.95}{2} = 0.975$, $c = u_{0.975} = 1.96$ (Tabelle 2)

3. $\Theta_u = 0.054 - 1.96 \cdot \sqrt{\frac{0.054 \cdot (1-0.054)}{500}} = 0.034$
   $\Theta_o = 0.054 + 1.96 \cdot \sqrt{\frac{0.054 \cdot (1-0.054)}{500}} = 0.074$
   p: $[0.034, 0.074]$

z.B. Gegeben: X normalverteilt, $n = 10$, $\bar{x} = 102$, $s^2 = 16$

Gesucht: Schätzintervall für μ und σ^2 mit Vertrauensniveau von 99 %

$n \le 30$ → μ gesucht → σ^2 unbekannt → Fall 2

1. $f = 10 - 1 = 9$, $p = \frac{1+0.99}{2} = 0.995$, $c = u_{0.995;9} = 3.25$ (Tabelle 4)

2. $\Theta_u = 102 - 3.25 \cdot \frac{\sqrt{16}}{\sqrt{10}} = 97.89$
   $\Theta_o = 102 + 3.25 \cdot \frac{\sqrt{16}}{\sqrt{10}} = 106.11$
   μ: $[97.89, 106.11]$

$n \le 30$ → σ^2 gesucht → Fall 3

1. $f = 10 - 1 = 9$
   $p_1 = \frac{1-0.99}{2} = 0.005$, $c_1 = u_{0.005;9} = 1.73$ (Tabelle 3)
   $p_2 = \frac{1+0.99}{2} = 0.995$, $c_2 = u_{0.995;9} = 23.59$ (Tabelle 3)

2. $\Theta_u = \frac{(10-1) \cdot 16}{23.59} = 6.1$
   $\Theta_o = \frac{(10-1) \cdot 16}{1.73} = 83.24$
   σ^2: $[6.1, 83.24]$

z.B. Gegeben: X normalverteilt, $n = 100$, $\bar{x} = 0.62$, $s = 0.35$

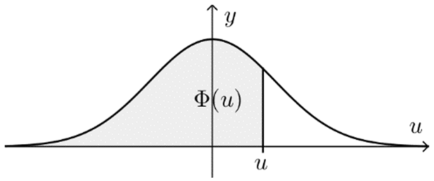
Gesucht: Schätzintervall für μ mit Irrtumswahrscheinlichkeit von 5 %

$n > 30$ → Fall 5 → μ gesucht → Fall 1 mit s als Schätzwert für σ

1. $\gamma = 1 - 0.05 = 0.95$, $p = \frac{1+0.95}{2} = 0.975$, $c = u_{0.975} = 1.96$ (Tabelle 2)

2. $\Theta_u = 0.62 - 1.96 \cdot \frac{0.35}{\sqrt{100}} = 0.613$
   $\Theta_o = 0.62 + 1.96 \cdot \frac{0.35}{\sqrt{100}} = 0.627$
   μ: $[0.613, 0.627]$

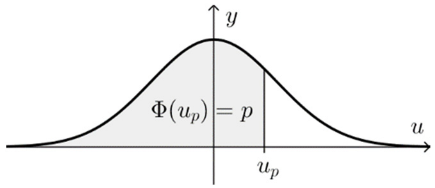
Tabelle 1: CDF $\Phi(u)$ der Standardnormalverteilung



$$P(U \le u) = \Phi(u)$$
$$P(U \ge u) = 1 - \Phi(u)$$
$$P(-u \le U \le u) = 2 \cdot \Phi(u) - 1$$
$$\Phi(-u) = 1 - \Phi(u)$$

u	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	0.8078	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
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	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
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

Tabelle 2: Quantile der Standardnormalverteilung

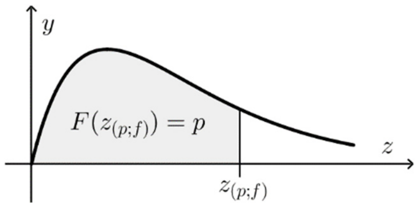


p : vorgegebene Wahrscheinlichkeit

u_p : zur Wahrscheinlichkeit p gehöriges Quantil

p	u_p	p	u_p
0.90	1.282	0.10	–1.282
0.95	1.645	0.05	–1.645
0.975	1.960	0.025	–1.960
0.99	2.326	0.01	–2.326
0.995	2.576	0.005	–2.576
0.999	3.090	0.001	–3.090

Tabelle 3: Quantile der Chi-Quadrat-Verteilung

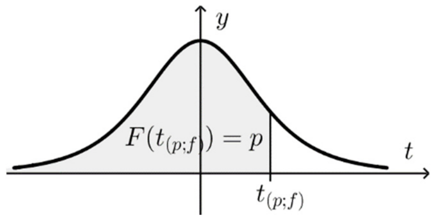


p : vorgegebene Wahrscheinlichkeit

$z_{(p;f)}$: zur Wahrscheinlichkeit p gehöriges Quantil bei f Freiheitsgraden

	p									
f	0.005	0.01	0.025	0.05	0.1	0.9	0.95	0.975	0.99	0.995
1	0.000	0.000	0.001	0.004	0.016	2.71	3.84	5.02	6.63	7.88
2	0.010	0.020	0.051	0.103	0.211	4.61	5.99	7.38	9.21	10.60
3	0.072	0.115	0.216	0.352	0.584	6.25	7.81	9.35	11.34	12.84
4	0.207	0.297	0.484	0.711	1.064	7.78	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	1.61	9.24	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	2.20	10.64	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09	21.95
9	1.73	2.09	2.70	3.33	4.17	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	17.28	19.68	21.92	24.72	26.76
12	3.07	3.57	4.40	5.23	6.30	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.26	7.26	8.55	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.86	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	28.41	31.41	34.17	37.57	40.00
22	8.6	9.5	11.0	12.3	14.0	30.8	33.9	36.8	40.3	42.8
24	9.9	10.9	12.4	13.8	15.7	33.2	36.4	39.4	43.0	45.6
26	11.2	12.2	13.8	15.4	17.3	35.6	38.9	41.9	45.6	48.3
28	12.5	13.6	15.3	16.9	18.9	37.9	41.3	44.5	48.3	51.0
30	13.8	15.0	16.8	18.5	20.6	40.3	43.8	47.0	50.9	53.7
40	20.7	22.2	24.4	26.5	29.1	51.8	55.8	59.3	63.7	66.8
50	28.0	29.7	32.4	34.8	37.7	63.2	67.5	71.4	76.2	79.5
60	35.5	37.5	40.5	43.2	46.5	74.4	79.1	83.3	88.4	92.0
70	43.3	45.4	48.8	51.7	55.3	85.5	90.5	95.0	100.4	104.2
80	51.2	53.5	57.2	60.4	64.3	96.6	101.9	106.6	112.3	116.3
90	59.2	61.8	65.6	69.1	73.3	107.6	113.1	118.1	124.1	128.3
100	67.3	70.1	74.2	77.9	82.4	118.5	124.3	129.6	135.8	140.2

Tabelle 4: Quantile der t-Verteilung von «Student»



p : vorgegebene Wahrscheinlichkeit

$t_{(p;f)}$: zur Wahrscheinlichkeit p gehöriges Quantil bei f Freiheitsgraden

	p				
f	0.9	0.95	0.975	0.99	0.995
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
22	1.321	1.717	2.074	2.508	2.819
24	1.318	1.711	2.064	2.492	2.797
26	1.315	1.706	2.056	2.479	2.779
28	1.313	1.701	2.048	2.467	2.763
30	1.310	1.697	2.042	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704
50	1.299	1.676	2.009	2.403	2.678
60	1.296	1.671	2.000	2.390	2.660
100	1.290	1.660	1.984	2.364	2.626
200	1.286	1.653	1.972	2.345	2.601
500	1.283	1.648	1.965	2.334	2.586
⋮	⋮	⋮	⋮	⋮	⋮
∞	1.282	1.645	1.960	2.326	2.576

$$\boxed{t_{(1-p;f)} = -t_{(p;f)}}$$