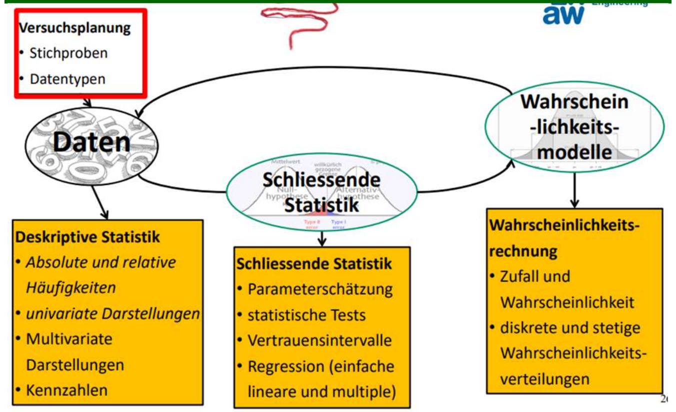
STOCHASTIK UND STATISTIK
URSULA MAYER



1. Deskriptive Statistik

Begriffe

***Merkmalsträger** bzw. statistische Einheiten:*

Objekte → interessierende Grössen beobachtet und erfasst werden (z.B., Menschen, Flugzeuge, …)

Grundgesamtheit

alle statistischen Einheiten, über die man Aussagen gewinnen möchte (z.B., alle Studierenden des Studiengangs Aviatik)

Vollerhebung

Bei jedem Individuum/ Objekt der Grundgesamtheit → Eigenschaften gemessen/abgefragt/erhoben

Stichprobe

Eigenschaften in Teilmenge der Grundgesamtheit erhoben. Sie kann zufällig, systematisch, willkürlich,…

Stichprobe (x_1,...,x_n) der Länge n

Merkmal** oder Variable → **kann man typisieren!

interessierende Grösse, die an den statistischen Einheiten in der Stichprobe beobachtet (gemessen, erhoben) wird. (z.B., Nationalität eines Passagiers, Beförderungsklasse eines Passagiers)

(Merkmals)Ausprägungen a_i

die verschiedenen Werte, die jedes Merkmal annehmen kann. Nationalität: Schweizer, Deutsche,…

Beispiel Partei

Eine solche Stichprobe (x_1,...,x_n) wäre zum Beispiel eine Folge von n Parteinamen x_i.

1.1 Merkmalstyp

Daten			
Qualitativ / Kategoriell ➢ sind keine Zahlenwerte im engeren Sinne ➢ haben nur endlich viele verschiedene Ausprägungen / Stufen Es wird eine Ausprägung und kein Ausmass angegeben. Insbesondere gibt es nur endlich viele Ausprägungen. **Stab-Diagramme, Balkendiagramme**		**Quantitativ / Metrisch** ➢ Merkmale, die numerische Werte annehmen Es wird ein Ausmadd angegeben. Das Ausmass wird mit Zahlen angegeben. **Stab-Diagramme, Balkendiagramme, Histogramme**	
Nominal Ausprägungen lassen sich nicht in natürlicher Weise anordnen ➢ Reine Katego-risierung ➢ Keine Ordnung	**Ordinal** Ausprägungen lassen sich in natürlicher Weise anordnen ➢ Ordnung vorhanden ➢ Rangierung möglich	**Diskret** Wertebereich ist abzählbar, aber möglicherweise unendlich gross	**Stetig** Merkmale, die prinzipiell jeden beliebigen Wert in einem Bereich annehmen können Alle ausprägungen in einem reelen intervall
Nationalität	Beförde-rungsklasse	Anzahl Passagiere	Geschwin-digkeit Flugzeug

Bsp: Würfel

Fragestellung	Ein Würfel mit den Zahlen 1 bis 6 wird viermal geworfen
Merkmalsausprägungen	Zahlen 1 bis 6
Messniveau	metrisch diskret

1.2 Häufigkeiten und Verteilungsfunktion

Begriffe:

Absolute und relative Häufigkeit

Für alle Datentypen bestimmbar

- ➢ Balkendiagramm bei kategoriellen und diskreten Daten

Kumulative (abs. und rel.) Häufigkeit

Für ordinale und quantitative Daten bestimmbar

Klassierte Stichproben

Sinnvoll bei metrisch stetigen Daten

- ➢ Einführung von absoluten, relativen und kumulativen Häufigkeiten für klassierte Stichproben
- ➢ Histogramm, PDF, CDF

Absolute und relative Häufigkeit

Grundgesamtheit Ω; Stichprobe der **Grösse / Stichprobenumfang** **(Objekteanzahl)** n mit Objekten $\omega_1, \omega_2, \dots, \omega_n$ und Merkmalen $X(\omega_1), X(\omega_2), \dots, X(\omega_n)$

Merkmale oft auch mit $X(w_i) = x_i$ gekennzeichnet

Merkmalsausprägungen: a_i $(i = 1, \dots, m)$

Beispiel Geburtsmonate

Aus einer Klasse Ω werden 3 Studierende $\omega_1, \omega_2, \omega_3$ ausgewählt und deren Geburtsmonate $X(\omega_1), X(\omega_2), X(\omega_3)$ notiert.

Die Funktion $X: \Omega \to \{1,2,3,\dots 12\}$ ordnet jedem Studierenden den Geburtsmonat zu.

Absolute Häufigkeit h_i

(Anzahl vom Wert (einer Merkmalsausprägung) a_i in der Urliste (der Ausprägung $a_1, a_2, \dots, a_m$))

Funktion: $h: \mathbb{R} \to \mathbb{R}$ alle nicht vorkommenden Werte = Null

Summe der Absoluten Häufigkeit = Stichproben Umfang

$$h_1 + h_2 + \dots + h_m = n$$

Relative Häufigkeit f_i ***von*** a_i *(empirische) Dichtefunktion*

$$\frac{Absolute\ Häufigkeit}{Stichprobenumfang} = \frac{h_i}{n} = f_i$$

Funktion $f\colon \mathbb{R} \to [0,1]$

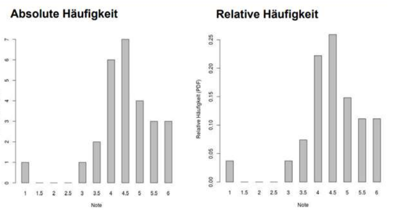
- **diskrete Merkmale**: probability mass function (**PMF**)
- **stetige Merkmale**: probability density function (**PDF**)

Summe der relative Häufigkeiten = 1

$$f_1 + f_2 + \ldots + f_m = 1$$

Beispiel: Noten von 27 Studierenden

Noten	Absolute Häufigkeit	Relative Häufigkeit
1	1	(1/27=) 3.7%
3	1	3.7%
3.5	2	7.4%
4	6	22.2%
4.5	7	25.9%
5	4	14.8%
5.5	3	11.1%
6	3	11.1%



Kumulative Häufigkeit / Kumulative Verteilfunktion CDF

Die **CDF** ist eine **Stufenfunktion**, die monoton von 0 bis 1 wächst, und an den Stellen a_i genau um f_i springt.

Anzahl Flugreisen a_i	1	2	3	4	5	Total
Absolute Häufigkeit h_i	9	8	5	7	1	30
Relative Häufigkeit f_i	9/30	8/30	5/30	7/30	1/30	1
Kumulative abs. Häufigkeit H_i	9	17	22	29	30	Summenwert
Kumulative rel. Häufigkeit F_i	9/30	17/30	22/30	29/30	30/30 = 1	

Stichprobe: $x_1, \ldots, x_n$

(empirische) **absolute Summenhäufigkeit** $H\colon \mathbb{R} \to \mathbb{R}$:

$$H(x) = Anzahl\ aller\ Stichprobenwerte \leq x\ bzw.$$

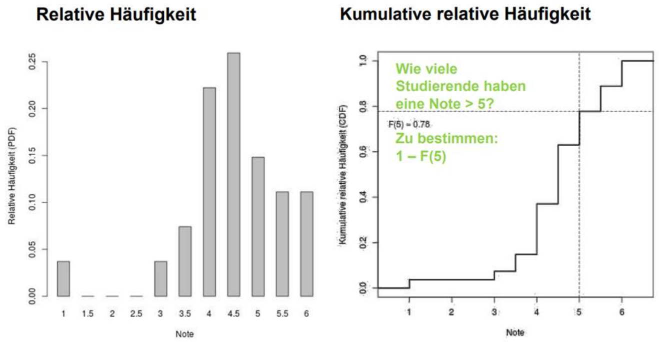
$$H(x) = \sum_{i: a_i \leq x} h_i$$

(empirische) **relative Summenhäufigkeit** $H\colon \mathbb{R} \to [0,1]$:

$$F(x) = \frac{H(x)}{n = Stichprobengrösse} = \sum_{i: a_i \leq x} f_i$$

Beispiel: Noten von 27 Studierenden

aufsummierte absolute/relative Häufigkeiten aller Beobachtungen, die kleiner oder gleich der jeweiligen Ausprägung sind.



1 – F(5) = 1 – 0.78 = 0.22

➢ **22% der Studierenden haben die Note 5.5 oder 6.**

1.3 Klassierte Stichproben

Absolute und relative Häufigkeit

grossen Stichproben ***metrisch stetiger Merkmale*** → Einteilung der Stichprobenwerte in ***Klassen*** → Merkmalsbereiche angrenzend (gleich gross nicht notwendig) Intervalle.

➢ Obere Intervallgrenzen zählen immer zum darauffolgenden Intervall.

➢ Einzelne Stichprobenwerte treten in den Hintergrund

➢ Es gibt nur noch Klassen und deren Intervalllänge

➢ die **absolute / relative Häufigkeit h_i / f_i** einer Klasse/eines Intervalls ergibt sich aus dem **Flächeninhalt** des darüber liegenden Rechtecks

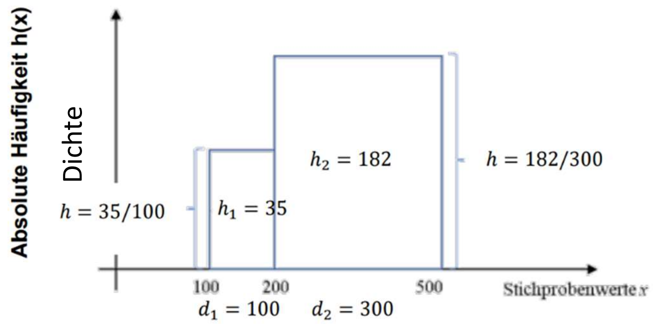
absolute / relative (Häufigkeits-) Dichte = h / f

Klassenbreite = d_i

$$h = \frac{h_i}{d_i} und\ f = \frac{f_i}{d_i}$$

Beispiel Ausgaben für Transport

Klasse c_i Von … bis weniger als …	[100,200[	[200,500[	[500,800[	[800,1000[	[1000,2000[	Total
Absolute Häufigkeit h_i des Intervalls	35	182	317	84	132	750
Relative Häufigkeit f_i des Intervalls	35/750	182/750	317/750	84/750	132/750	1
Säulenbreite im Histogramm d_i	100	300	300	200	1000	

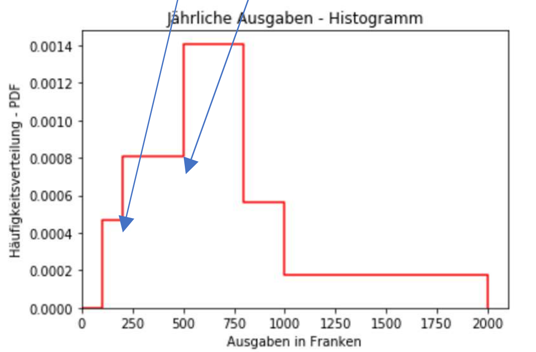


Relative Häufigkeitsdichtefunktion (PDF)

$$f(x) = \frac{absolute\ Häufigkeitsdichtefunk\ldots}{Stichprobengrösse} = \frac{h(x)}{n}$$

Beispiel über Klassen

Klasse c_i Von … bis weniger als …	[100,200[	[200,500[	[500,800[	[800,1000[	[1000,2000[	Total
Absolute Häufigkeit h_i des Intervalls	35	182	317	84	132	750
Relative Häufigkeit f_i des Intervalls	35/750	182/750	317/750	84/750	132/750	1
Säulenbreite im Histogramm d_i	100	300	300	200	1000	
Säulenhöhe im Histogramm (abs. Häufigkeits-dichte h: R → R)	35/100	182/300	317/300	84/200	132/1000	
Säulenhöhe im Histogramm (rel. Häufigkeits-dichte f: R → [0,1])	(35/750) /100	(182/750) /300	(317/750) /300	(84/750) /200	(132/750) /1000	



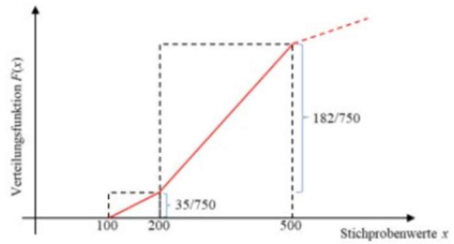
Kumulative Verteilungsfunktion (CDF)

Integration der relativen Häufigkeitsdichtefunktion (PDF) $f(x)$

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

Beispiel Klassen

Klasse a_i Von … bis weniger als …	[100,200[	[200,500[	[500,800[	[800,1000[	[1000,2000[	Tota
Relative Häufigkeit f_i des Intervalls	35/750	182/750	317/750	84/750	132/750	1
Werte der kum. Verteilungsfunktion (CDF) $F(x)$ an der rechten Intervallgrenze	35/750	217/750	534/750	618/750	1	



1.4 Kenngrössen

Lagemasse	Streuungsmasse	Schiefemasse
Zentrum	Breite	Asymneetrie
beschreiben um welchen „mittleren“ Wert die Daten verteilt sin	geben an wie „breit“ die Verteilung ist, d.h. wie stark die Werte um das Zentrum streuen	beschreiben die Form der Verteilung

Definition Lagemasse

Arithmetisches Mittel (auch empirischer Mittelwert)

- Durchschnitt der Daten

Median

- teilt eine Stichprobe in zwei (gleich grosse) Hälften
- entspricht dem 2. Quartil

Modus (oder Modalwert)

- Ist der häufigste Wert, der in der Stichprobe vorkommt

Arithmetisches Mittel (Mittelwert / Durchschnitt)

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

m = Anzahl der verschiedenen Merkmale

$n = \sum_{i=1}^{m} h_i$

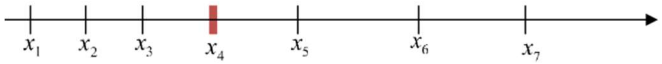
Median (liegt in der Mitte)

→ Daten der Grösse nach ordnen (klein nach gross)

Ungerade Anzahl von Date (n):

$$\tilde{x} = mittlere\ Wert\ der\ geordneten\ Daten = x_{\left[\frac{n+1}{2}\right]}$$

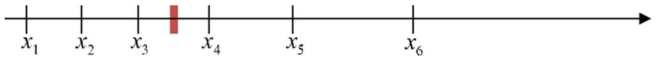
$$Median(x_1, \dots, x_7) = x_{[4]}$$



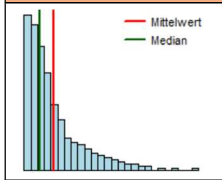
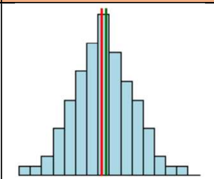
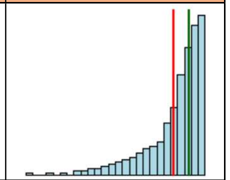
Gerade Anzahl von Daten (n):

$$\tilde{x} = \frac{1}{2}\cdot\left(x_{\left[\frac{n}{2}\right]} + x_{\left[\frac{n}{2}+1\right]}\right)$$

$$Median(x_1, \dots, x_6) = \frac{1}{2}(x_{[3]} + x_{[4]})$$



Median oder Mittelwert

Rechtsschief	Symmetrisch	Linksschief
		
$x_{mod} < x_{med} < \bar{x}$	$x_{mod} = x_{med} = \bar{x}$	$x_{mod} > x_{med} > \bar{x}$

Weitere «Mittelwerte»

*Harmonisches Mittel **h** von **n** Zahlen **$x_1, x_2, \dots, x_n$** :*

$$h = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}}$$

Das ist das Mittel von Quotienten (Ratios, Verhältnisse, ect.)

*Geometrische Mittel **g** von **n** Zahlen **$x_1, x_2, \dots, x_n$** :*

$$g = x_1 \cdot x_2 \cdot \ldots \cdot x_n^{\frac{1}{n}}$$

Das ist das Mittel der Wachstumsraten (Zinsrechnungen)

Beispiel Gewinnspiel – empirisch

Resultat	Sonstiges	3er Pasch	Full House	Grosse Strasse	4er Pasch	5er Pasch
Gewinn a_i	-1	0	4	8	10	20
Absolute Häufigkeit h_i	74	13	3	5	4	1
Relative Häufigkeit f_i	0.74	0.13	0.03	0.05	0.04	0.01
Kum. abs. Häufigkeit H_i	74	87	90	95	99	100
Kum. rel. Häufigkeit F_i	0.74	0.87	0.90	0.95	0.99	1

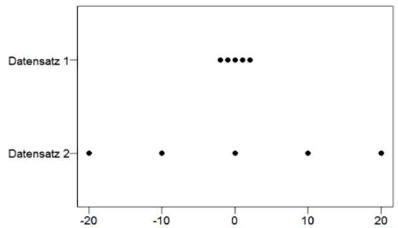
Arithmetische Mittel:

$$\bar{x} = \frac{1}{100}(-1\cdot 74 + 0\cdot 13 + 4\cdot 3 + 8\cdot 5 + 10\cdot 4 + 20\cdot 1) = \frac{38}{100} = 0.38\ [CHF]$$

Median	Modus
$Median(x_1, \dots x_n) = x_{med} = -1$	$Modus(x_1, \dots x_n) = x_{mod} = -1$

Definition Streuungsmasse

Datensatz 1: -2 -1 0 1 2 Datensatz 2: -20 -10 0 10 20



→ Arithmetisches Mittel/Median für beide 0

→ Streuung im zweiten Datensatz grösser

Varianz

mittlere quadratische Abweichung der Beobachtungen vom arithmetischen Mittel

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

$$s^2 = \frac{1}{n}\sum_{i=1}^{m} h_i(a_i - \bar{x})^2 = \left(\sum_{i=1}^{m} a_i^2 \cdot f_i\right) - \bar{x}^2$$

n = Stichprobengrösse / m = Anzahl Merkmale

Korrigierte Varianz

$$S_{kor}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \quad = \quad \frac{n}{n-1}s^2$$

Standartabweichung

Quadratwurzel der Varianz

- Vorteil: gleiche Einheit wie die Beobachtungen
- Nachteil: Standardabweichung (wie auch Varianz) nicht robust gegen Ausreisser.

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

Korrigierte Standartabweichung

$$s_{kor} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} \quad = \quad \sqrt{\frac{n}{n-1}} \cdot s$$

Kovarianz (Erweiterung Varianz auf zwei Variablen)

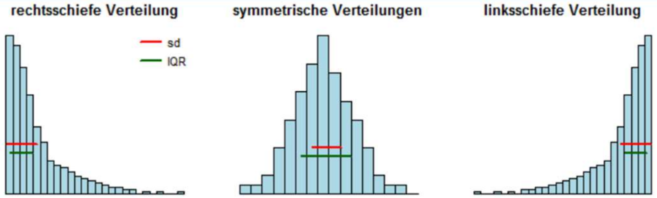
$$s_{xy} = \overline{xy} - \bar{x} \cdot \bar{y}$$

Alternative Streuungsmasse

Interquartile Range (IQR) (Interquantilsabstand)

deckt die zentralen 50% der Daten ab

$$IQR = Q_3 - Q_1 = R_{q_3} - R_{q_1}$$



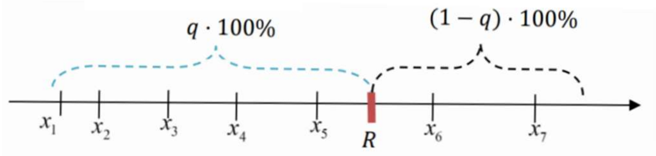
Mean absolute deviation (MAD)

- Robustes Gegenstück zur Standardabweichung
- Mittelwertbildungen durch Mediane ersetzen

$$MAD_x = 1.4826 \cdot median(|x_1 - \tilde{x}|, |x_2 - \tilde{x}|, \ldots, |x_n - \tilde{x}|)$$

Perzentil / Quantil

Eine Zahl R heisst q-Perzentil / Quantil, falls R die Stichprobenwerte im Verhältnis ***q:(1-q)*** teilt



0.25-Quantil	0.5-Quantil	0.75-Quantil
1. Quantil	**2. Quantil**	**3. Quantil**

Berechnung:

$n \cdot q$ ganze Zahl	$R_q = \frac{1}{2} \cdot (x_{n \cdot q} + x_{n \cdot q+1})$
$n \cdot q$ Bruch	$R_q = x_{n \cdot q}$ (*nächst grösser Zahl*)

Beispiele Quantile

0.25-Quantil:
$n \cdot q = 7 \cdot 0.25 = 1.75$ mit $\lceil n \cdot q \rceil = 2$
$x_{\lceil n \cdot q \rceil} = x_{[2]} = 1$ (2. Stichprobenwert)
$R_{0.25} = 1$

0.5-Quantil:
$n \cdot q = 7 \cdot 0.5 = 3.5$ mit der nächstgrösseren ganzen Zahl $\lceil n \cdot q \rceil = 4$
$x_{\lceil n \cdot q \rceil} = x_{[4]} = 3$ (4. Stichprobenwert)
$R_{0.5} = 3$

0.75-Quantil:
$n \cdot q = 7 \cdot 0.75 = 5.25$ mit $\lceil n \cdot q \rceil = 6$
$x_{\lceil n \cdot q \rceil} = x_{[6]} = 4$ (6. Stichprobenwert)
$R_{0.75} = 4$

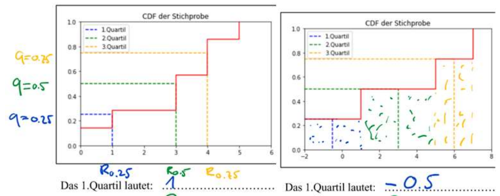
Quantile bei klassierten Daten

$$R_q = F^{-1}(x)$$

sucht Klasse $[a, b[$ mit $F(a) \le q \le F(b)$

$$R_q = a + \frac{(b-a) \cdot (q - F(a))}{F(b) - F(a)}$$

Beispiel Quantile aus und über CDF



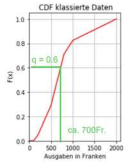
Das 1.Quartil lautet: 1	Das 1.Quartil lautet: -0.5
Der Median lautet: 3	Der Median lautet: 3
Das 3. Quartil lautet: 4	Das 3. Quartil lautet: 6

Beispiel: Ausgabe für transport

q = 0.6

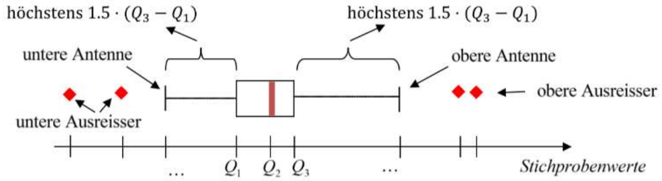
Klasse, so dass $F(a) \le 0.6 \le F(b)$:

- Klasse [500,800[
- Formel: $R_q = a + \frac{(b-a)(q-F(a))}{F(b)-F(a)} = 500 + \frac{(800-500)(0.6-0.29)}{0.71-0.29} = 720.5Fr.$



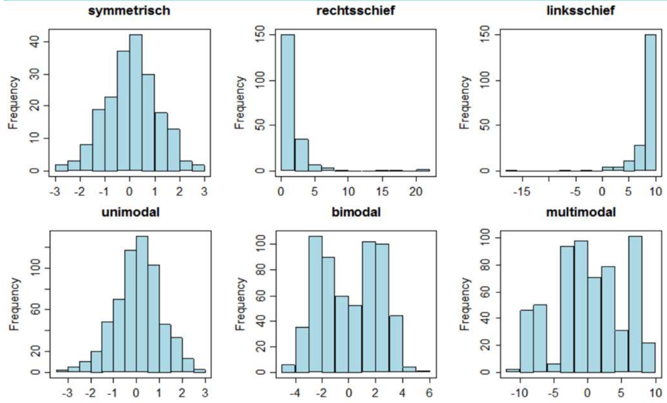
Klasse a_i Von … bis weniger als …	[100,200[	[200,500[	[500,800[	[800,1000[	[1000,2000[	Total
Relative Häufigkeit f_i des Intervalls	35/750	182/750	317/750	84/750	132/750	1
Werte der kum. Verteilungsfunktion (*CDF*) *F*(*x*) an der rechten Intervallgrenze	35/750 =0.05	217/750 = 0.29	534/750 = 0.71	618/750 = 0.82	1	

Boxplot



→ nicht geeignet bei bimodalen Verteilungen

Form der Verteilung in einem Histogramm



2. Deskriptive Statistik (mehrere Merkmale)

Darstellung von Variabeln

Kategorielle Variablen

- Häufigkeitstabelle → mässig Übersicht, langweilig
- Modus → beschränkte Information
- Balkendiagramm → generell gut geeignet

Metrische Variablen

- Mittelwert/Standardabweichung → bei symmetrischer Verteilung
- Median/IQR/Quantile → robuste Kennzahlen
- Histogramm → generell gut geeignet
- Boxplot → auch gut, aber zeigt wenig Details

2.1 Bivariate Daten

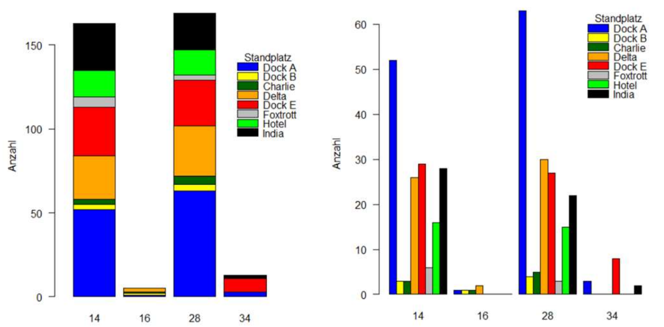
1. Kategoriell vs. Kategoriell

Informationen zu den landenden Flugzeugen eines Tages am Flughafen Zürich.
Wie könnte man Landebahnen und Standplätze gleichzeitig darstellen?

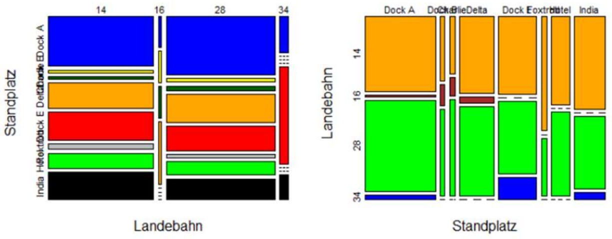
Kontingenztabelle

Standplatz\ Landebahn	Dock A	Dock B	Charlie	Delta	Dock E	Foxtrott	Hotel	India	Total
14	52	3	3	26	29	6	16	28	163
16	1	1	1	2	0	0	0	0	5
28	63	4	5	30	27	3	15	22	169
34	3	0	0	0	8	0	0	0	13
Total	119	8	9	58	64	9	31	52	350

Gruppierte Balkendiagramme



Mosaik-Plot

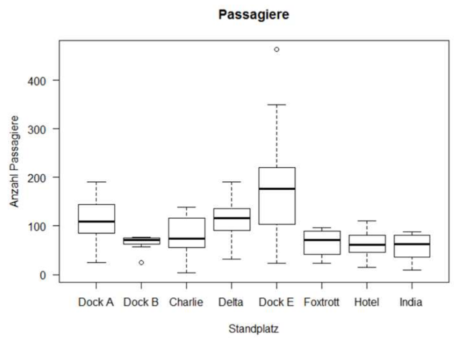


2. Metrisch vs. Kategoriell

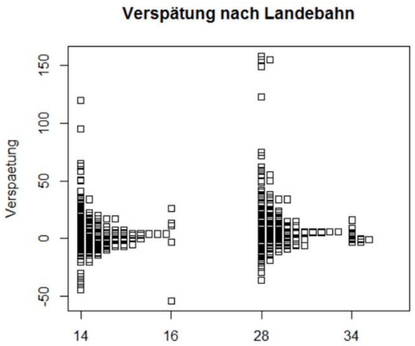
Tabelle mit Kennzahlen

Standplatz	Mittelwert Passagiere	Standardabweichung Passagiere
Dock A	112.6	37.1
Dock B	65.0	17.4
Charlie	79.1	43.5
Delta	114.2	33.4
Dock E	163.5	79.5
Foxtrott	65.3	27.3
Hotel	61.3	25.6
India	58.4	24.1

Boxplots

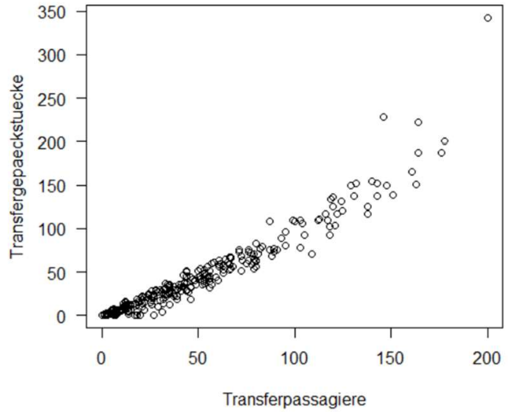


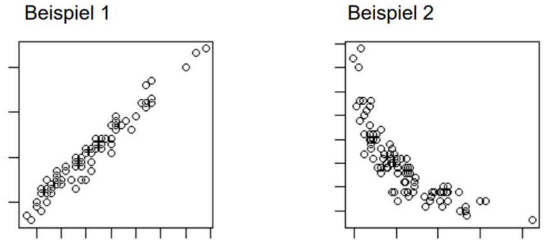
Stripcharts



3. Metrisch vs. Metrisch

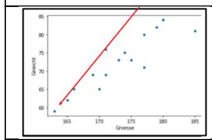
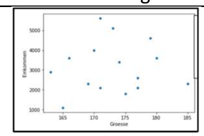
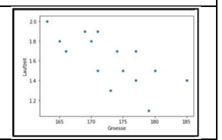
Scatterplot bzw. Streudiagramm





Zusammenhang		
	Beispiel 1	Beispiel 2
Form	«linearer»	monoton abfallend
Richtung	positiv	negativ
Stärke	stark	mittel

Beispiele Scatterplot (Streudiagramm)

«je grösser umso schwerer»	Die Punkte liegen ziemlich regellos verteilt. Man erkennt keinen Zusammenhang.	Mit grösser werdenden x-Werten werden die y-Werte kleiner.
		
Linear, positiv, stark	Nicht-linear, schwach	Linear, negativ, stark

Darstellungsmöglichkeiten für bivariate Daten

Kategoriell vs. Kategoriell:

- Kreuztabelle → wenig Übersicht, langweilig
- Gruppierte Balkendiagramme → mässig übersichtlich
- Mosaikplot → gut, aber wenig verbreitet

Metrisch vs. Kategoriell:

- Tabelle mit Mittelwerten → beschränkte Information, langweilig
- Boxplots → generell gut geeignet
- Stripcharts → auch gut, wenn "stacked"

Metrisch vs. Metrisch:

- Korrelation → Mass für linearen Zusammenhang
- Scatterplot → zeigt die Verteilung der Daten

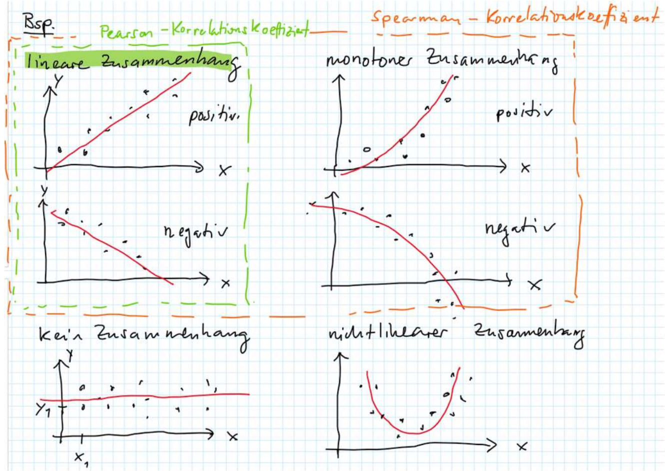
Korrelation

Pearson-Korrelation:

Misst die Stärke des linearen Zusammenhangs

Spearman-Korrelation, a.k.a. Rang-Korrelation:

Misst die Stärke des monotonen Zusammenhangs



Pearson Korrelation (Korrelationskoeffizient)

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})\cdot(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2\cdot\sum_{i=1}^{n}(y_i-\bar{y})^2}}$$

$$= \frac{\frac{1}{n}\cdot\sum_{i=1}^{n}(x_i-\bar{x})\cdot(y_i-\bar{y})}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}\cdot\sqrt{\frac{1}{n}\cdot\sum_{i=1}^{n}(y_i-\bar{y})^2}} = \frac{s_{xy}}{s_x\cdot s_y}$$

$$r_{xy} = \frac{s_{xy}}{s_x\cdot s_y} = \frac{\overline{xy}-\bar{x}\cdot\bar{y}}{\sqrt{\overline{x^2}-\bar{x}^2}\cdot\sqrt{\overline{y^2}-\bar{y}^2}}$$ (am einfachsten zu verwenden)

$\bar{x} = \frac{1}{n}\cdot\sum_{i=1}^{n} x_i$	$\bar{y} = \frac{1}{n}\cdot\sum_{i=1}^{n} y_i$
$\overline{xy} = \frac{1}{n}\cdot\sum_{i=1}^{n} x_i y_i$	$\overline{x^2} = \frac{1}{n}\cdot\sum_{i=1}^{n} x_i^2$

- Es gilt immer: $-1 \le r_{xy} \le 1$
- $r_{xy} > 0$ → positive Korrelation («je mehr desto mehr»)
- $r_{xy} < 0$ → negative Korrelation («je mehr desto weniger»)
- $r_{xy} = 1$ → Punkte exakt auf Gerade mit positiver Steigung
- $r_{xy} = -1$ → Punkte exakt auf Gerade mit negativer Steigung
- r_{xy} *nahe bei* ± 1 → Punkte streuen eng um die Gerade
- r_{xy} ist nicht die Steigung der Geraden
- $r_{xy} = 0$ → kein linearer Zusammenhang (anderer Zusammenhang kann bestehen)
- Korrelation nie ohne Blick auf den Scatterplot beurteilen

Beispiel Empirische Kenngrössen

$x = [1,2,3]$ und $y = [4,-1,2]$

Die arithmetischen Mittel lauten:

$\bar{x} = \frac{1}{3}[1+2+3] = 2$

$\bar{y} = \frac{1}{3}[4-1+2] = \frac{5}{3}$

Die Varianzen lauten:

$s_x{}^2 = \frac{1}{3}[(1-2)^2+(2-2)^2+(3-2)^2] = \frac{2}{3}$

$s_y{}^2 = \frac{1}{3}[(4-\frac{5}{3})^2+(-1-\frac{5}{3})^2+(2-\frac{5}{3})^2] = \frac{1}{3}[(\frac{7}{3})^2+(\frac{8}{3})^2+(\frac{1}{3})^2] = \frac{38}{9}$

Die Standardabweichungen lauten:

$s_x = \sqrt{\frac{2}{3}}$ $\qquad s_y = \frac{\sqrt{38}}{3}$

Die Kovarianz lautet:

$s_{xy} = \frac{1}{3}[(1-2)(4-\frac{5}{3})+(2-2)(-1-\frac{5}{3})+(3-2)(2-\frac{5}{3}) = \frac{-2}{3}$

Der Korrelationskoeffizient lautet:

$r_{xy} = \frac{s_{xy}}{s_x s_y} = -\sqrt{\frac{3}{19}}$

Spearman – Rangkorrelation

Rangkorrelation Pearson Korrelation vorzuziehen:

- **Zusammenhang nicht linear** sondern **monoton**
- es **Ausreisser** geben könnte
- die Werte (xi, yi) nicht glockenförmig verteilt

$$r_{x_R y_R} = \frac{\sum_{i=1}^{n}\left(\left(rang(x_i)-\frac{n+1}{2}\right)\cdot\left(rang(y_i)-\frac{n+1}{2}\right)\right)}{\sqrt{\sum_{i=1}^{n}\left(rang(x_i)-\frac{n+1}{2}\right)^2}\cdot\sqrt{\sum_{i=1}^{n}\left(rang(y_i)-\frac{n+1}{2}\right)^2}}$$

$$mit\ \frac{n+1}{2} = \overline{rg(x)} = \overline{rg(y)}$$

Einzelne Teile einzeln berechnen, da Taschenrechner schnell überfordert wird!!

Es kann vorkommen, dass zwei oder mehr Werte für x denselben Wert annehmen. In diesem Fall wird diesen verbundenen Rängen der ***Durchschnittsrang*** zugewiesen.

$$r_{x_R y_R} = r_{Sp} = 1 - \frac{\left(6\cdot\sum_{i=1}^{n} d_i^2\right)}{n(n^2-1)}$$

$d_i = rg(x_i) - rg(y_i)$ = **Differenz der Ränge**

Diese Formel ist nur anwendbar, wenn alle Ränge unterschiedlich sind.

Beispiel Spearman-Korrelation

i	1	2	3	4	5	6	
x_i	59	35	43	23	42	27	
y_i	14.6	11.8	14.3	13.0	14.2	11.0	
$rg(x_i)$	6	3	5	1	4	2	$\overline{rg(x)} = 3.5$
$rg(x_i) - \overline{rg(x)}$	2.5	-0.5	1.5	-2.5	0.5	-1.5	
$rg(y_i)$	6	2	5	3	4	1	$\overline{rg(y)} = 3.5$
$rg(y_i) - \overline{rg(y)}$	2.5	-1.5	1.5	-0.5	0.5	-2.5	

Beispiel Verbundene Ränge

i	1	2	3	4	5	6
x_i	23	27	35	35	42	59
$rg(x_i)$	1	2	(3+4)/2 =3.5	(3+4)/2 =3.5	5	6

Korrelation Interpretation

- Korrelationen mit kleinem Absolutbetrag können rein zufällig von 0 verschieden sein (ab 0,7 kann man eine «normale» Interpretation beginnen)
- Vorsicht bei Stichprobengrösse; bei zwei Punkten ist r_{xy} immer ± 1
- Ein statistisch gesicherter und im Streudiagramm offensichtlicher Zusammenhang noch lange kein kausaler (ursächlicher) Zusammenhang
  - X kann von Y oder Y von X abhängen
  - Eine dritte Grösse Z (Confounding factor) kann X und Y beeinflussen
  - Der Zusammenhang kann völlig zufällig sein

Confounding Factor

Übersehen eines dritten Merkmales, das mit den anderen inhaltlich in Zusammenhang steht.

Scheinkorrelation (oder von einer Scheinkausalität)

Beispiel Storch

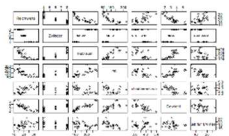
Mit der hohen Korrelation zwischen der Zahl der Störche und der Zahl der Geburten zwischen 1900 und 1970 könnte man statistisch beweisen, dass der Klapperstorch die Babys bringt…

2.2 Mehrere Merkmale

Multivariate Beobachtungen: Simpson Paradoxon

- Tritt häufig auf, wenn Daten heterogener Gruppen aggregiert werden
- Falscher Schluss aus einfachen Balkendiagrammen, da Gruppengrösse nicht berücksichtigt wird
- Beinhaltet mindestens 3 Variablen (Beispiel von Hochschulen)
  - Zielvariable (Zulassung)
  - Beobachtete Variable (Geschlecht)
  - Störvariable (Departemente)

Scatterplot Matrix →



3. Kombinatorik

Fakultät

$n! = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 3 \cdot 2 \cdot 1$ *0!=1*

Binomialkoeffizient

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$$

$$\binom{n}{0} = \frac{n!}{0! \cdot (n-0)!} = \frac{n!}{0! \cdot n!} = \frac{n!}{1 \cdot n!} = 1$$

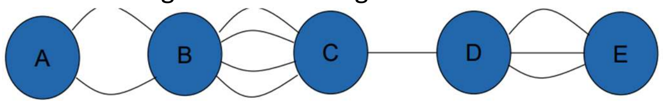
Symmetrie	Rekursion
$\binom{n}{k} = \binom{n}{n-k}$	$\binom{n+1}{k+1} = \binom{n}{k} + \binom{n}{k+1}$

Beispiel Lotto 6 aus 45

$$\binom{45}{6} = \frac{45!}{6!\,(45-6)!} = \frac{45 \cdot 44 \cdot 43 \cdot 42 \cdot 41 \cdot 40 \cdot 39!}{6! \cdot 39!}$$
$$= \frac{45 \cdot 44 \cdot 43 \cdot 42 \cdot 41 \cdot 40}{6!} = \frac{5'864'443'200}{720} = 8'145'060$$

Produkteregel

Wie viele Wege von A nach E gibt es?



$z = n_1 \cdot n_2 \cdot \ldots \cdot n_k = 2 \cdot 4 \cdot 1 \cdot 3 = 24$

		Ohne Wiederholung bzw. Zurücklegen	Mit Wiederholung bzw. Zurücklegen
Permutation	Auswahl: $k = n$ Reihenfolge wichtig	Permutation ohne Wiederholung $z = n!$	Permutation mit Wiederholung $z = \frac{n!}{n_1! \cdot \ldots \cdot n_s!}$
Variation	Auswahl: k Reihenfolge wichtig	Geordnete Stichprobe ohne Zurücklegen $(k < n)$ $z = \frac{n!}{(n-k)!}$	Geordnete Stichprobe mit Zurücklegen $(k \text{ ist beliebig})$ $z = n^k$
Kombination	Auswahl: k Reihenfolge unwichtig	Ungeordnete Stichprobe ohne Zurücklegen $(k < n)$ $z = \binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}$	Ungeordnete Stichprobe mit Zurücklegen $(k \text{ ist beliebig})$ $z = \binom{n+k-1}{k}^*$

$$* \; z = \binom{n+k-1}{k} \rightarrow \binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}$$

4. Elementare Wahrscheinlichkeitsrechnung

4.1 Diskrete Wahrscheinlichkeitsräume

Zufallsexperiment

Ein Versuch bzw. eine Situation, in welchem/r das Ergebnis nicht deterministisch (bei gleicher Eingabe immer die gleiche Ausgabe) vorbestimmt ist.

→ Bei einer Wiederholung immer exakt dasselbe Resultat einstellen würde?

Nein → Definition für ein Zufallsexperiment erfüllt!

Ergebnis

Zufallsexperiments.

- 1 x Würfelwurf: 2 (jedes Ergebnis hat Wahrscheinlichkeit 1/6 bzw. 100/6 %)

Fairer Würfel → Idealisierung → jede Seite gleiche

→ Bei jedem Experiment genau ein Ergebnis!

Ergebnisraum / Wahrscheinlichkeitsraum

Menge aller möglichen Ergebnisse.

$$\boldsymbol{Ergebnisraum}\ \Omega = \{1,2,\ldots,6\}$$

$$\boldsymbol{Mächtigkeit}\ (\boldsymbol{Anzahl\ der\ Elemente}) = |\Omega|$$

Ereignis

Konstellation, die unter Umständen von mehreren Ergebnissen erfüllt wird.

Teilmenge A von Ω

Bsp.: 1 x Würfeln: A = „alle geraden Zahlen“ = {2,4,6}

Teilmenge kann eine Wahrscheinlichkeit zugeordnet werden.

Spezielle Ereignisse

- **Elementarereignisse**: {ω1},{ω2}, … enthalten nur ein Ergebnis

  Bsp.: 1 x Würfeln: ω1 = {1}, ω2 = {2}, ω3 = {3}, ω4 = {4}, ω5 = {5} und ω6 = {6}

- **Gegenereignis** von A: A^c mit A ∪ A^c = Ω

  Bsp.: von Fussballmatch → A = «Winti gewinnt»: A^c = «Winti verliert oder es gibt ein Unentschieden»

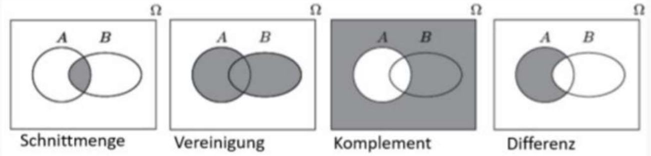
- **Sicheres Ereignis**: A = Ω

  Bsp.: vom Treibstoffverbrauch → A = “Es wird mehr als 0t verbraucht.”

- **Unmögliches Ereignis**: A = Ø (leere Menge)

Mengenoperationen

- Schnittmenge: $A \cap B$
- Vereinigung: $A \cup B$
- Komplement: A^C
- Differenz: $A \setminus B$ (A ohne B)



Beispiele Mengenoperationen

- Treibstoffverbrauch: A = [35t, 37t], B = [36t, 40t]

  Schnittmenge: [36t, 37t]

  Vereinigung: [35t,40t]

- Würfeln: A = «Gerade Zahl» und B = «Zahl kleiner gleich 4»

  Differenz A \ B = «Es wird eine 6 gewürfelt»

  Komplement AC: «Ungerade Zahl»

Ereignisraum

Menge aller Teilmengen / Potenzmenge von Ω

$$\boldsymbol{Ereignisraum} = 2^{|\Omega|}$$

Beispiel fairer Würfel Ω = {1,2,3,4,5,6}:

Teilmengen: {}, {1}, … , {6}, {1,2}, {1,3}, …, {1,2,3,4,5,6}

Anzahl aller Teilmengen von {1,2,3,4,5,6}: $\mathbf{2}^{|\Omega|} = 2^6 = 64$

Berechnung von Wahrscheinlichkeiten

Ereignis nicht mit Sicherheit **voraussagbar**

→ Angabe von **Wahrscheinlichkeiten**

Wahrscheinlichkeit für Ereignis A: P(A)

Modellbasiert: Laplace-Wahrscheinlichkeit

Laplace-Raum ≙ die Ergebnisse sind gleich wahrscheinlich

$$Laplace - Formel \rightarrow P(A) = \frac{|A|}{|\Omega|}$$

$$= \frac{Anzahl\ \text{"}günstige\text{"}\ Fälle}{Anzahl\ \text{"}mögliche\ Fälle} = \frac{Elemente\ von\ A}{Ergebinsraum}$$

Beispiel fairer Würfel

A: Wahrscheinlichkeit eine gerade Zahl wird gewürfelt. (Teilmenge von Ergebnisraum)

$$A = \{2,4,6\} \subset \Omega = \{1,2,\ldots,6\}$$

$$P(A = geraden\ Zahlen) = \frac{|\{2,4,6\}|}{|\{1,2,3,4,5,6\}|} = \frac{3}{6} = 0.5$$

B: Wahrscheinlichkeit eine Zahl ≥ 3 wird gewürfelt.

$$B = \{3,4,5,6\} \subset \Omega = \{1,2,\ldots,6\}$$

$$P(B = Zahl\ \geq 3) = \frac{|\{3,4,5,6\}|}{|\{1,2,3,4,5,6\}|} = \frac{4}{6} = \frac{2}{3}$$

Datenbasiert: Frequentistische Wahrscheinlichkeit

kein Modell/Theorie verfügbar → **Wahrscheinlichkeit aufgrund vergangener Beobachtungen**

$P(A) = r_A$ r_A = relative Häufigkeit des Ereignisses A

Beispiel Passagierzahlen

- Gesucht: P(A = „nicht mehr als 200 Passagiere“)
- Wahrscheinlichkeit wird geschätzt als Anteil der früheren Flüge, bei welchen nicht mehr als 200 Passagiere erschienen sind.
- Schätzung aus der Vergangenheit:
  - Bei insgesamt 100 Flügen sind bei 13 Flügen mehr als 200 Passagiere erschienen:

    → P(A) = 1 – 0.13 = 0.87

Subjektive Wahrscheinlichkeit

Wahrscheinlichkeit = Mass des persönlichen Glaubens (Bayesianische Interpretation)

Experimentell nicht wiederholbar

Beispiel:	Beispiel Fussballspiel:
• Atomarer Störfall • Flugzeugabsturz	• P (A=„Winterthur gewinnt“) • Subjektive Einschätzung → P(A) = 30%

Zähldichte (PMF, PDF)

mit welcher Wahrscheinlichkeit die möglichen Ergebnisse des Zufallsexperiments auftreten.

$$\varrho: \Omega \to [0,1],$$
$$\sum_{\omega \in \Omega} \varrho(\omega) = 1$$

Beispiel Würfel:

$$\varrho: \{1,2,3,4,5,6\} \to [0,1],$$
$$\varrho(\omega) = \frac{1}{6} \text{ für alle } \omega \in \Omega$$

Wahrscheinlichkeitsmass P (Wahrscheinlichkeitsraum)

Gibt an, mit welcher Wahrscheinlichkeit das Ereignis, das durch die Menge M beschrieben wird, eintritt.

$$P: 2^{\Omega} \to [0,1], P(M) = \sum_{\omega \in M} \varrho(\omega), M \subseteq \Omega$$

Beispiel Würfel:

A: Wahrscheinlichkeit eine gerade Zahl wird gewürfelt. (Teilmenge von Ergebnisraum)

$$P(A) = \sum_{\omega \in A} \varrho(\omega) = \varrho(2) + \varrho(4) + \varrho(6) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$$

Axiome von Kolmogorov

- Axiom 1: $0 \leq P(A) \leq 1$
- Axiom 2: $P(\Omega) = 1$ (sicheres Ereignis)
- Axiom 3: Falls $A \cap B = \emptyset$, dann gilt $P(A \cup B) = P(A) + P(B)$

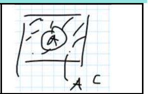
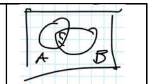
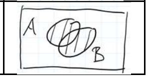
Eigenschaften der Wahrscheinlichkeit (Diskreter Wahrscheinlichkeitsraum)

Jeder diskrete Wahrscheinlichkeitsraum (Ω, P) hat die folgenden Eigenschaften:

(A1) (Unmögliches Ereignis) $P(\{\}) = 0$
(A2) (Sicheres Ereignis) $P(\Omega) = 1$
(A3) (Komplementäres Ereignis) $P(\Omega \backslash A) = 1 - P(A)$
(A4) (Vereinigung) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
(A5) (Sigma-Additivität) $P(A_1 \cup A_2 \cup A_3 \cup \ldots) = P(A_1) + P(A_2) + P(A_3) + \ldots$ falls die Ereignisse $A_1, A_2, A_3, \ldots$ paarweise disjunkt sind.

Ist (Ω, P) ein Laplace-Raum, so gilt: $P(M) = \frac{|M|}{|\Omega|}$.

Wahrscheinlichkeit eines Ereignisses

Nicht gerade	$\neg A \triangleq A^c = \{1,3,5\}$	
Gerade und ≥ 3	$A \wedge B \triangleq A \cap B = \{4,6\}$	
Gerade oder ≥ 3	$A \vee B \triangleq A \cup B = \{2,3,4,5,6\}$	

4.2 Zufallsvariablen

Eine Grösse, die unter bestimmten, konstanten Bedingungen unterschiedliche, durch den Zufall bedingte Werte annehmen kann.

Zufallsvariable X ordnet jedem Element des **Ergebnisraumes Ω** eine reelle Zahl zu.

$$X: \Omega \to \mathbb{R}, \qquad X(\omega), \omega \in \Omega$$

Zufallsvariablen → Grossbuchstaben (X, Y)
Realisation → Kleinbuchstaben (x, y)

Beispiel Würfel

X = gewürfelte Augenzahl
Realisierung von X
x = 3, wenn die Augenzahl 3 gewürfelt wird.
x = 4, wenn die Augenzahl 4 gewürfelt wird.

***Diskrete** Zufallsvariable*

- X nimmt endlich oder unendlich abzählbar viele Werte an.
- Überall dort, wo gezählt wird
- Beispiel: Passagiere, Flugzeuge, Unfälle, etc.

***Stetige** Zufallsvariable*

- Wenn X jeden beliebigen Wert eines Intervalls annehmen kann.
- Überall, wo gemessen wird
- Beispiel: Länge, Gewicht, Temperatur, etc.

Beispiel: Augensumme beim Würfeln mit zwei fairen Würfeln

Ergebnisraum:

$$\Omega = \{1, \ldots, 6\} \times \{1, \ldots, 6\}$$
$$= \{(1,1), (1,2), \ldots, (2,1), \ldots (6,6)\}$$

Anzahl Ergebnisse (Mächtigkeit):

$$|\Omega| = 6 \cdot 6 = 36 \; Ergebnisse$$

Laplace-Raum ≙ Ergebnisse gleich wahrscheinlich

PMF:

$$\varrho(\omega) = \varrho(6,6) = \frac{1}{36}$$

Augensumme: Funktion auf Ω ≙ **Zufallsvariablen**

$$X: \Omega \to \{2,3,\ldots,12\}$$
$$X(\omega) = \omega_1 + \omega_2 \; mit \; \omega = (\omega_1, \omega_2) \in \Omega$$

Ergebnisse lassen sich «elegant» ausdrücken:

$$\{x = 2\} = \{\omega \in \Omega | X(\omega) = 2\} = \{(1,1)\}$$
$$\{x > 5\} = \{\omega \in \Omega | X(\omega) > 5\}$$

Weitere Beispiele mit PMF und CDF

	x	2	3	4	5	6	7	8	9	10	11	12
	Ergebnisse aus Ω	(1,1)	(1,2) (2,1)	(1,3) (2,2) (3,1)	(1,4) (2,3) (3,2) (4,1)	(1,5) (2,4) (3,3) (4,2) (5,1)	(1,6) (2,5) (3,4) (4,3) (5,2) (6,1)	(2,6) (3,5) (4,4) (5,3) (6,2)	(3,6) (4,5) (5,4) (6,3)	(4,6) (5,5) (6,4)	(5,6) (6,5)	(6,6)
PMF	$f(x) = P(X = x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$
CDF	$F(x) = P(X \leq x)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{15}{36}$	$\frac{21}{36}$	$\frac{26}{36}$	$\frac{30}{36}$	$\frac{33}{36}$	$\frac{35}{36}$	1

Dichtefunktion PMF

$$p_i = f(x_i) = P(X = x_i)$$

$f(x_i)$ = Wahrscheinlichkeit bei x_i

- Die Summe aller $p_i = 1$, also $\sum_i p_i = 1$
- Wenn alle p_i gleich sind, heisst die Wahrscheinlichkeitsfunktion diskrete Gleichverteilung

$$\boldsymbol{f: \mathbb{R} \to [0,1], f(x) = P(X = x)}$$

Beispiele PMF

Wahrscheinlichkeit Augensumme 2 würfeln

$$P(X = 2) = P(\{(1,1)\}) = \varrho((1,1)) = \frac{1}{36}$$

Wahrscheinlichkeit Augensumme 3 würfeln

$$P(X = 3) = P(\{(1,1), (2,1)\}) = \varrho((1,1)) + \varrho((2,1))$$
$$= \frac{2}{36}$$

$f(2) = \frac{1}{36}$ $\qquad f(2.2) = 0$ $\qquad f(-101) = 0$

Kumulative Verteilungsfunktion CDF

Aufsummieren aller Wahrscheinlichkeiten

- Für $F(x)$ gilt stets $f(-\infty) = 0$ und $F(\infty) = 1$
- Eine kumulative Verteilungsfunktion ist monoton wachsend, d.h. für $u < v$ gilt $F(u) \leq F(v)$
- Für diskrete Zufallsvariablen ist $F(x)$ eine Treppenfunktion

$$F: \mathbb{R} \to [0,1], \boldsymbol{F(x) = P(X \leq x)}$$

Wahrscheinlichkeit Augensumme grösser als 5 würfeln

$$P(X \leq 4) = F(4) = \frac{6}{36} = \frac{1}{6}$$

(P{(1,1),(1,2),(1,3),(2,1),(2,2),(3,1)})

Wahrscheinlichkeit Augensumme zwischen 7 und 10

$$P(7 \leq X \leq 10) = P(X \leq 10) - P(X < 7)$$
$$= F(10) - F(6) = \frac{33}{36} - \frac{15}{36} = \frac{18}{36} = \frac{1}{2}$$

4.3 Kenngrössen

Diskrete Wahrscheinlichkeitsraum (Ω, P)

Zufallsvariable X: $\Omega \to \mathbb{R}$.

→ Kenngrössen von X folgende Eigenschaften:

- Linearität des Erwartungswertes
  $E(X+Y) = E(X) + E(Y)$ *und* $E(\alpha \cdot X) = \alpha \cdot E(X)$ *mit* $\alpha \in \mathbb{R}$
- Verschiebungssatz für die Varianz
  $V(X) = E(X^2) - \left(E(X)\right)^2 = \left(\sum_{x\in\mathbb{R}} P(X=x) \cdot x^2\right) - \left(E(X)\right)^2$
- $V(\alpha \cdot X + \beta) = \alpha^2 \cdot V(X)$ *mit* $\alpha, \beta \in \mathbb{R}$

Erwartungswert einer Zufallsvariable (Durchschnitt, Lagemass)

$$E(X) = \mu = \sum_i x_i \cdot P(X = x_i)$$
$$= \sum_i x_i \cdot p(x_i) = \sum_i x_i \cdot p_i$$

Beispiel Würfelwurf

$$E(X) = \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \frac{1}{6} \cdot 3 + \frac{1}{6} \cdot 4 + \frac{1}{6} \cdot 5 + \frac{1}{6} \cdot 6 = 3.5$$

Empirische Varianz einer Zufallsvariablen (Streuungsmass)

$$V(X) = Var(X) = E\left[\left(X - E(X)\right)^2\right]$$
$$= \sum_{i=1}^{N} \left(x_i - E(X)\right)^2 \cdot p_i = \boldsymbol{\sigma^2}$$

Verschiebungssatz:

$$\boldsymbol{\sigma^2} = V(X) = Var(X) = E(X^2) - [E(X)]^2$$
$$= \left(\sum_{i=1}^{N} (x_i)^2 \cdot p_i\right) - \left(E(X)\right)^2$$

Beispiel Würfelwurf

$$Var(X) = (1-3.5)^2 \cdot \frac{1}{6} + (2-3.5)^2 \cdot \frac{1}{6} + (3-3.5)^2 \cdot \frac{1}{6} + (4-3.5)^2 \cdot \frac{1}{6} + (5-3.5)^2 \cdot \frac{1}{6} + (6-3.5)^2 \cdot \frac{1}{6}$$
$$= (-2.5)^2 \cdot \frac{1}{6} + (-1.5)^2 \cdot \frac{1}{6} + (-0.5)^2 \cdot \frac{1}{6} + (0.5)^2 \cdot \frac{1}{6} + (1.5)^2 \cdot \frac{1}{6} + (2.5)^2 \cdot \frac{1}{6} = 2.92$$

Alternative:

$$E(X^2) = \frac{1}{6} \cdot 1^2 + \frac{1}{6} \cdot 2^2 + \frac{1}{6} \cdot 3^2 + \frac{1}{6} \cdot 4^2 + \frac{1}{6} \cdot 5^2 + \frac{1}{6} \cdot 6^2 = 15.17$$

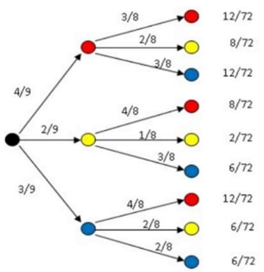
$$Var(X) = 15.17 - 3.5^2 = 2.92$$

Standartabweichung

$$\sigma = \sqrt{\sigma^2}$$

4.4 Bedingte Wahrscheinlichkeiten

Ereignisbäume *(Kugelbeispiel)*



- ➢ Aufteilung der Äste gemäss Wahrscheinlichkeiten
- ➢ Wahrscheinlichkeit am Ende eines Pfades **Multiplikation** Wahrscheinlichkeiten entlang Pfades
- ➢ Wahrscheinlichkeit mehrere Pfade beschriebenes Ereignis **Summe** aller Blätter
- ➢ **Endwahrscheinlichkeiten** aller Blätter addieren ist gleich 1

	ELISA +	ELISA -	Total
HIV +	4'985	15	5'000
HIV -	14'925	980'075	995'000
Total	19'910	980'090	1'000'000

Begriffe

Prävalenz Wahrscheinlichkeit, dass eine zufällig herausgegriffene Blutspende HIV-positiv ist

$$P(HIV^+) = \frac{5'000}{1'000'000} = 0.5\%$$

Sensitivität Anteil der HIV-positiven Blutproben, welche durch den Test bekannt werden

$$P(E^+|HIV^+) = \frac{4'985}{5'000} = 99.7\%$$

Spezifität Umgekehrte Frage, wie gross der Anteil an HIV-negativen Personen ist, die auch ein negatives Testresultat erhalten

$$P(E^-|HIV^-) = \frac{980'075}{995'000} = 98.5\%$$

Definition Bedingte Wahrscheindlichkeit

Bedingte Wahrscheinlichkeit von B unter der Bedingung A.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{günstige\ Fälle}{möglich\ \ Fälle} \qquad \cap = und$$

Beispiele Raucher

	F (Frauen)	M (Männer)	Summe
R (Raucher/in)	$P(F \cap R) = \frac{100}{1400}$	$P(M \cap R) = \frac{400}{1400}$	$P(R) = \frac{500}{1400}$
N (Nichtraucher/in)	$P(F \cap N) = \frac{200}{1400}$	$P(M \cap N) = \frac{700}{1400}$	$P(N) = \frac{900}{1400}$
Summe	$P(F) = \frac{300}{1400}$	$P(M) = \frac{1100}{1400}$	$P(\Omega) = \frac{1400}{1400} = 1$

Wahrscheinlichkeit einen Raucher/in anzutreffen.

$$P(R) = \frac{500}{1400} = \frac{5}{14} = 35.7\%$$

Sie treffen eine Frau Wahrscheinlichkeit das sie raucht.

$$P(R|F) = \frac{100}{300} = 33.3\%$$

Satz von Bayes

	B	$\bar{B}$	
A	$P(A \cap B)$	$P(A \cap \bar{B})$	$P(A)$
$\bar{A}$	$P(\bar{A} \cap B)$	$P(\bar{A} \cap \bar{B})$	$P(\bar{A})$
	$P(B)$	$P(\bar{B})$	1

1 In der Kontingenztabelle:

Beschreibt die Beziehung zwischen den zwei bedingten Wahrscheinlichkeiten $P(A|B)$ und $P(B|A)$ und erlaubt das «Umkehren» von bedingten Wahrscheinlichkeiten.

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)}$$

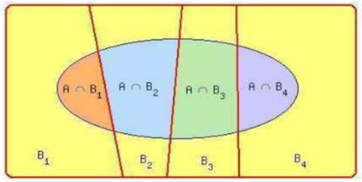
$$P(A|B) \cdot P(B) = P(B|A) \cdot P(A) = P(A \cap B)$$

Satz der totalen Wahrscheinlichkeit

$B_1, B_2, \ldots, B_k$ sind paarweise disjunkte Ereignisse, und

$B_1 \cup B_2 \cup \cdots \cup B_k = \Omega$

d.h. genau ein B_i tritt immer ein.



B_i ergeben eine disjunkte Zerlegung des W-Raums Ω.
Jedes Element von einem weiteren Ereignis A liegt so genau in einem Bi , es gilt:

$$P(A) = \sum_{i=1}^{k} P(A|B_i) \cdot P(B_i)$$

Beispiel Flugzeug Pünktlichkeit bei Niederschlag

Bei Niederschlag beträgt die Wahrscheinlichkeit, dass ein Flugzeug pünktlich ist 50%. Bei sonnigem Wetter ist es in 90% der Fälle pünktlich. Morgen wird es mit Wahrscheinlichkeit 70% sonnig sein. Wie gross ist die Wahrscheinlichkeit, dass morgen ein Flugzeug pünktlich eintreffen wird?

$P(pünktlich) = P(pünktlich|schön) \cdot P(schön) + P(pünktlich|schlecht) \cdot P(schlecht) = 0.9 \cdot 0.7 + 0.5 \cdot 0.3 = 0.78 = 78\%$

Satz von Bayes (allgemein)

$P(A) = \sum_{i=1}^{k} P(A|B_i) \cdot P(B_i)$ in $P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)}$ einsetzen

$$P(B_i|A) = \frac{P(B_i \cap A)}{P(A)} = \frac{P(A|B_i) \cdot P(B_i)}{\sum_{i=1}^{k} P(A|B_i) \cdot P(B_i)}$$

4.5 Stochastische Unabhängigkeit

Zwei Ereignisse A und B sind **unabhängig**, wenn sie sich gegenseitig nicht beeinflussen:

$$P(A|B) = P(A) \quad mit \quad P(B) \neq 0$$

Für unabhängige Ereignisse gilt:

$$P(A \cap B) = P(A) \cdot P(B)$$

Zwei Zufallsvariablen sind **unabhängig** wenn:

$$X: \Omega \to \mathbb{R} \qquad Y: \Omega \to \mathbb{R}$$

$$P(X = x \wedge Y = y) = P(X = x) \cdot P(Y = y)$$
$$für\ alle\ x, y\ \in \mathbb{R}$$

$P(X = x \wedge Y = y)$

→Wahrscheinlichkeit Ereignisse gleichzeitig eintreffen. Zufallsvariable X nimmt wert x an und Zufallsvariable Y nimmt gleichzeitig den Wert y an.

gemeinsame Verteilung (Verbundsverteilung)

$$f(x, y) = P(X = x \wedge Y = y)$$

Satz (Stochastische Unabhängigkeit – Ereignisse)

Diskrete Wahrscheinlichkeit → (Ω, P)

Ereignisse → $A\ und\ B \subseteq \Omega$

Folgende Eigenschaften sind äquivalent:

i. $A\ und\ B$ sind stochastisch unabhängig.
ii. $A\ und\ \Omega/B = \bar{B}$ sind stochastisch unabhängig.
iii. $\Omega/A = \bar{A}\ und\ B$ sind stochastisch unabhängig.

5. Spezielle Verteilungen

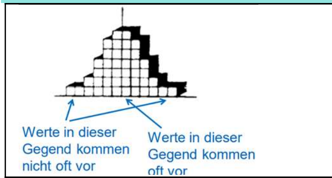
5.1 Diskrete und stetige Verteilung

diskreten Zufallsvariable: Lücken zwischen den Werten; **zählt** etwas und kann entsprechend nur ganzzahlige Werte annehmen.

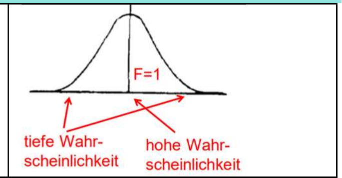
stetige Zufallsvariable: kontinuierliches Spektrum von möglichen Werten; Werte durch **Messung**

diskret	stetig
Anzahl Studierende, die an einem bestimmten Tag anwesend sind	Temperatur an einem bestimmten Ort zu einer bestimmten Zeit
Anzahl Tore in einem bestimmten Fussballspiel	Höhe eines Heissluftballons zu einem bestimmten Zeitpunkt

Wahrscheinlichkeitsdichte: Mass für die beobachtete relative Häufigkeit



Histogramm Summe der relativen Häufigkeiten beträgt 1



Wahrscheinlichkeitsdichte Fläche unter der Kurve beträgt 1

Wahrscheinlichkeitsdichtefunktion (PDF)

$f: \mathbb{R} \to \mathbb{R}$ $\quad f(x) \geq 0\ für\ alle\ x \in \mathbb{R}$

$\int_{-\infty}^{\infty} f(u)du = 1$

Die Werte der Dichtefunktion $f(x)$ beschreiben keine Wahrscheinlichkeit!

- $f(x) \neq P(X = k)$ das gilt nur im diskreten Fall!
- $f(x)$ kann grösser sein als 1s

Kennzahlen stetiger Zufallsvariablen

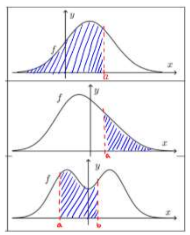
Erwartungswert $\quad \mu = E(X) = \int_{-\infty}^{\infty} f(x) \cdot x\, dx$

Varianz $\quad \sigma^2 = V(X) = \int_{-\infty}^{\infty} f(x) \cdot \left(x - E(X)\right)^2 dx$

Standardabweichung $\quad \sigma = \sqrt{V(X)}$

Bei stetigen Zufallsv. gilt: Für alle $x_0 \in \mathbb{R}$ ist $P(X = x_0) = 0$

Berechnung von Wahrscheinlichkeiten



$\boldsymbol{P(X \leq a)} = F(a) = \int_{-\infty}^{a} f(x)dx$

$\boldsymbol{P(X > a)} = 1 - P(X \leq a) = 1 - F(a)$

$= 1 - \int_{-\infty}^{a} f(x)dx = \int_{a}^{\infty} f(x)dx$

$\boldsymbol{P(a < X \leq b)} = P(X \leq b) - P(X \leq a)$

$= F(b) - F(a) = \int_{a}^{b} f(x)dx$

$\boldsymbol{P(a \leq X \leq b)} = P(a \leq X < b)$

$= P(a < X < b) = P(a < X \leq b)$

	diskrete Zufallsvariablen	stetige Zufallsvariablen
Dichtefunktion / PMF bzw. PDF	$f(x) = P(X = x)$	$f(x) = F'(x) \neq P(X = x)$!!!
Kumulative Verteilungsfunktion/ CDF	$F(x) = P(X \leq x) = \sum_{x \leq X} f(x)$	$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)du$
Wahrscheinlichkeiten	$P(a \leq X \leq b) = \sum_{a \leq x \leq b} f(x)$	$P(a \leq X \leq b) = \int_{a}^{b} f(x)dx$
Graphische Darstellung von f	Stabdiagramm	Graph
Erwartungswert	$E(X) = \sum_{x \in \mathbb{R}} f(x) \cdot x$	$E(X) = \int_{-\infty}^{\infty} f(x) \cdot x\, dx$
Varianz	$V(X) = \sum_{x \in \mathbb{R}} f(x) \cdot (x - E(X))^2$	$V(X) = \int_{-\infty}^{\infty} f(x) \cdot (x - E(X))^2\, dx$

Zufallsvariable	**diskret**	**stetig**
Wertebereich	abzählbar (z.B. Zähldaten)	beliebig (z.B. Messung)
Wahrscheinlichkeits- bzw. **Dichtefunktion**	▪ $p_i = P(X = x_i)$ ▪ $\sum_{i=1}^{N} p_i = 1$	$f(x)$ Eigenschaften: ▪ $f(x) \geq 0$ ▪ stetig ▪ $\int_{-\infty}^{\infty} f(z)dz = 1$
Kumulative Verteilungsfunktion $F(x) = P(X \leq x)$	▪ $F(x) = \sum_i p_i$	▪ $F(x) = \int_{-\infty}^{x} f(z)dz$
Kennzahlen: ▪ **Erwartungswert** $\mu = E[X]$ ▪ **Varianz** $\sigma^2 = Var(X)$ $(= E[(X - E[X])^2])$	▪ $\mu = \sum_i x_i p_i$ ▪ $\sigma^2 = \sum_i (x_i - \mu)^2 p_i$	▪ $\mu = \int_{-\infty}^{\infty} x f(x)dx$ ▪ $\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)dx$

5.2 Diskrete Verteilungen

Diskrete Wahrscheinlichkeitsfunktion

Wahrscheinlichkeitsfunktion (PMF) einer diskreten Zufallsvariablen gibt zu jedem möglichen Wert xi die entsprechende Wahrscheinlichkeit $\boldsymbol{pi = f(xi)}$ an.

X	x_1	x_2	x_3	...	x_N
$f(x)$	p_1	p_2	p_3	...	p_N

Notation $\quad p_i = f(x_i) = P(X = x_i) \quad f: \mathbb{R} \rightarrow [0,1]$

$$\sum_i p_i = 1$$

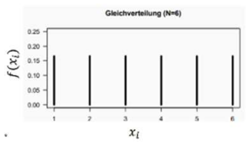
Diskrete kumulative Verteilungsfunktion

Zu jeder diskreten Wahrscheinlichkeitsfunktion a **kumulative Verteilungsfunktion** (CDF): $F(x) = P(X \leq x)$ gibt Wahrscheinlichkeit an Zufallsvariable X einen Wert annimmt, der kleiner oder gleich vorgegebenen Wert

- Für $F(x)$ gilt stets $F(-\infty) = 0$ und $F(\infty) = 1$.
- CDF ist monoton wachsend, d.h. für $u < v$ gilt $F(u) \leq F(v)$.
- Für diskrete Zufallsvariablen $F(x)$ Treppenfunkt.

→ **4.3 Kenngrössen**

Uniformverteilung / Gleichverteilung



Der Ergebnisraum ist
$\Omega = [a, (a+1), \ldots, (b-1), b\}$
Der Ergebnisraum ist endlich

Erwartungswert $\quad E(X) = \frac{a+b}{2}$

Varianz $\quad V(X) = \frac{(b-a+2)\cdot(b-a)}{12} = \frac{(b-a+1)^2-1}{12}$

Notation $\quad X \sim Unif(\{a, b\})$

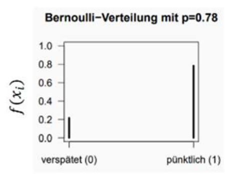
Beispiel Würfelwurf

$\Omega = \{1, 2, 3, 4, 5, 6\}$, $|\Omega| = 6$ und X = Augenzahl

X	$x_1 = 1$	$x_2 = 2$	$x_3 = 3$	$x_4 = 4$	$x_5 = 5$	$x_6 = 6$
$p(\cdot)$	$p_1 = 1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$

Bernoulli-Verteilung

Experiment mit zwei möglichen Ausgängen



Erfolg	Misserfolg
Ja	Nein
1	0

X	$x_1 = 0$	$x_2 = 1$
$f(.)$	$1 - p$	p

$p = P(X = 1) = Erfolgswahrscheinlichkeit$

Erwartungswert $\quad E(X) = 0 \cdot (1-p) + 1 \cdot p = p$

Varianz $\quad V(X) = (0-p)^2 \cdot (1-p) + (1-p)^2 \cdot p = p \cdot (1-p)$

Den Parameter $\boldsymbol{p}$ *kann man aus vergangenen Daten schätzen (falls man ihn nicht kennt)*

Notation $\quad X \sim Bernoulli(p)$

Beispiel Swiss-Flug

Ein Swiss-Flug ist pünktlich Im Jahr 2018 waren 109'748 von 140'074 Flügen pünktlich:

$$\hat{p} = \frac{109748}{140074} = 78.35\%$$

Binominalverteilung

beschreibt n unabhängige Bernoulli Experimente mit konstanter Erfolgswahrscheinlichkeit p, die n-mal nacheinander ausgeführt wurden.

Die Anzahl Erfolge X hat dann die folgende Verteilung:

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

$n = Anzahl\ von\ Bernoulli\ Experimente$
$p = Wahrscheinlichkeit\ für\ Erfolgreiches\ Ereignis$
→ n und p fix gegeben

$p^k(1-p)^{n-k}$	Wahrscheinlichkeit für ein Ereignis mit k Erfolgen und n – k Misserfolgen.
$\binom{n}{k}$	Anzahl Ereignisse mit k Erfolgen in n Versuchen

Der Ergebnisraum ist $\quad \Omega = \{a, (a+1), \ldots, (b-1), b\}$
Der Ergebnisraum ist endlich

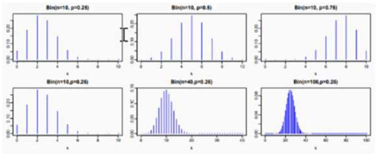
Erwartungswert $E(S) = E(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} p = n \cdot p$

Varianz $V(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} V(X_i) = n \cdot p \cdot (1-p)$

Notation $X \sim Bin(n,p)$

kann auch als Summe S von n unabhängigen Wiederholungen einer Bernoulli-verteilten Zufallsgrösse Xi mit Parameter p beschrieben werden

$$S = \sum_{i=1}^{n} X_i$$



Beispiel Businessklasse

In der Businessklasse sitzen generell 70% Männer. Ein Flugzeug hat 30 Plätze in der Businessklasse. Alle sind besetzt. Wie viele Frauen sind dabei?

X = Anzahl Frauen in der Businessklasse

n = 30; p = 1 - 0.7 $X \sim \mathrm{Bin}(30, 0.3)$

$P(X = k) = \binom{30}{k} 0.3^k (0.7)^{30-k}$

Beispiel Würfel

Es wird 12x gewürfelt. Was ist die Wahrscheinlichkeit genau dreimal eine Sechs zu werfen?

X = Anzahl 6er in 12 Versuchen zu werfen

$$X \sim Bin(n,p) = X \sim Bin\left(12, \frac{1}{6}\right)$$

$$P(X = 3) = \binom{12}{3} \cdot \left(\frac{1}{6}\right)^3 \cdot \left(1 - \frac{1}{6}\right)^{12-3} = 0.1974$$

Poissonverteilung

Für die Wahrscheinlichkeit für das Eintreten einer bestimmten Anzahl gleichartiger Ereignisse.

X = "Anzahl Ereignisse pro Zeiteinheit/Gebiet»

Unfälle in einer Fabrik oder Strassen	Das Eintreffen von Kunden an einem Schalter

Wahrscheinlichkeitsverteilung für Anzahl Ereignisse X:

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

$\lambda = Erwartungswert$ $k = max.Anzahl$

$$P(X > k) = 1 - P(X \leq k) = 1 - e^{-\lambda} \cdot \sum_{k=0}^{k} \frac{\lambda^k}{k!}$$

λ = Rate, mit welcher die Ereignisse in vorgegebenen Zeiteinheit oder Gebiet eintreffen

Notation: $X \sim Pois(\lambda)$

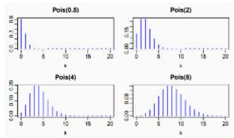
Ergebnisraum: $\Omega = \{0,1,2,3,4,\ldots\}$

keine klar definierte Obergrenze

Erwartungswert $E(X) = \lambda$

Varianz $V(X) = \sigma^2 = \lambda$

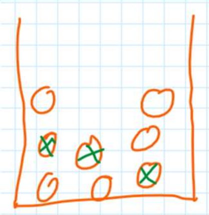
σ = Standardabweichung
Für kleine λ stark rechtsschief;
Je grösser λ, desto symmetrischer;
Generell gute Symmetrie ab $\lambda > 10$



Binominal vs Poisson

Binomial	Poisson
• bekannten Anzahl von n Einzelversuchen, d.h. Wertebereich von X: $\{0, \ldots, n\}$	• Grösse der Population ist unbekannt, d.h. Wertebereich von X hat keine klar definierte Obergrenze.
• Erfolgswahrscheinlichkeit p für den Einzelversuch	• Rate λ, mit welcher das Ereignis auftritt

Hypergeometrische Verteilung



N Kugeln

M Kugeln markiert

Entnahme einer *Stichprobe* des Umfangs *n*

X «Anzahl der markierten Kugeln in der Stichprobe»

$$PMF: P(X = k) = \frac{\binom{M}{k} \cdot \binom{N-M}{n-k}}{\binom{N}{n}}$$

M = Total von Objekten einer Sorte
N = Anzahl Objekte in einer Urne
n = Anzahl gezogener Objekte ohne Zurücklegen
k = Anzahl der Durchführungen

$N > M$ *und* $N > n$ *macht sinn für* $k = min(n, M)$

Ergebnisraum $\Omega = \{0,1,2,3,4,\ldots,N\}$ *endlich*

Erwartungswert $E(X) = n \cdot \frac{M}{N} = n \cdot (relative\ Häufigkeit)$

Varianz $V(X) = n \cdot \frac{M}{N} \cdot \left(1 - \frac{M}{N}\right) \cdot \frac{N-n}{N-1}$

Notation: $X \sim H(N, M, k)$

Beispiel Lotto 6 aus 49

M = 6; N = 49; n = 6; k = 4

$$P(X = 4) = \frac{\binom{6}{4} \cdot \binom{43}{2}}{\binom{49}{6}} \rightarrow exakt\ 4\ korrekt$$

Wahrscheinlichkeit für mindestens 4 richtige Zahlen

$$P(X \geq 4) = P(X = 4) + P(X = 5) + P(X = 6)$$

$$= \frac{\binom{6}{4} \cdot \binom{43}{2} + \binom{6}{5} \cdot \binom{43}{1} + \binom{6}{6} \cdot \binom{43}{0}}{\binom{49}{6}}$$

$$= \sum_{k=4}^{6} P(X = k) = 9.8714 \cdot 10^{-4}$$

Zusammenfassung

	Hypergeometrische Vert.	Binomialverteilung	Poissonverteilung
Dichtefunktion für x=0,1,.. Sonst 0	$P(X=x) = \frac{\binom{M}{x} \cdot \binom{N-M}{n-x}}{\binom{N}{n}}$	$P(X=x) = \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x}$	$P(X=x) = \frac{\lambda^x}{x!} e^{-\lambda}$
Erwartungs-wert	$n \cdot \frac{M}{N}$	$n \cdot p$	λ
Varianz	$n \cdot \frac{M}{N} \cdot (1 - \frac{M}{N}) \cdot \frac{N-n}{N-1}$	$n \cdot p \cdot (1-p)$	λ
Standardab-weichung	$\sqrt{n \cdot \frac{M}{N} \cdot (1 - \frac{M}{N}) \cdot \frac{N-n}{N-1}}$	$\sqrt{n \cdot p \cdot (1-p)}$	$\sqrt{\lambda}$
Parameter	n, N, M	n, p	λ

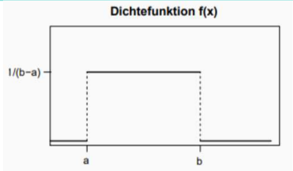
Verteilung	W'keitsfunktion	$E[X]$ und $Var(X)$	Anwendung
Gleichverteilung	$P(X=k)=p$ für alle k; $k \in 1,\dots,n$	$\frac{n+1}{2}$ $\frac{n^2-1}{12}$	Alle Ereignisse sind gleich wahrscheinlich
Bernoulli $X \sim$ Bernoulli(p)	$P(X=1)=p$, $P(X=0)=1-p$ R: `dbinom(x, size = 1, prob = p)`	p $p(1-p)$	Erfolg eines Versuchs.
Binomial $X \sim$ Bin(n, p)	$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$ R: `dbinom(x, size = n, prob = p)`	np $np(1-p)$	Anzahl Erfolge in n unab. Bernoulli-Versuchen.
Poisson $X \sim$ Pois(λ)	$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}$ R: `dpois(x, lambda = λ)`	λ λ	Anzahl Ereignisse mit konstanter Eintreffrate λ.

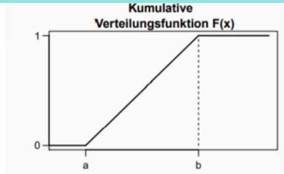
5.3 Stetige Verteilungen

Stetige Zufallsvariable

- Zufallsvariable X kann jeden beliebigen Wert eines Intervalls annehmen.
- Anwendung überall dort wo gemessen wird
- Beispiel: Länge, Gewicht, Temperatur, Zeit,...

Stetige Uniformverteilung / Gleichverteilung (Intervall)





$$f(x)=\begin{cases}0 & \text{falls } x<a\\ 1/(b-a) & \text{falls } a\le x\le b\\ 0 & \text{falls } x>b\end{cases} \qquad F(x)=\begin{cases}0 & \text{falls } x<a\\ (x-a)/(b-a) & \text{falls } a\le x\le b\\ 1 & \text{falls } x>b\end{cases}$$

Erwartungswert $E(X)=\frac{a+b}{2}$

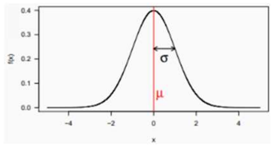
entspricht der Mitte des Intervalls [a,b]

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

Varianz erhöht sich mit grösserer Intervalllänge

Notation $X\sim Unif([a,b])$

Normalverteilung



Stetige Zufallsvariable X folgt Normalverteilung mit den Parametern $\mu,\sigma \in \mathbb{R}, \sigma>0$, wenn Dichtefunktion PDF:

$$\varphi_{\mu,\sigma}(x)=\frac{1}{\sigma\cdot\sqrt{2\cdot\pi}}\cdot e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$$

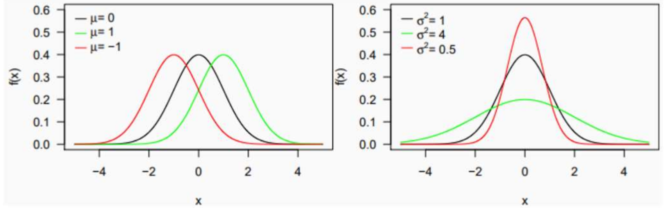
$\mu=n\cdot p$ und $\sigma^2=n\cdot p\cdot q=n\cdot p\cdot(1-p)$

$\mu=Erwartungswert=E(X)$

$\sigma^2=Varianz=V(X)$

Notation: $X\sim N(\mu,\sigma)$

- Die Dichtefunktion ist symmetrisch um $x=\mu$.
- Auf beiden Seiten asymptotisch Annäherung an 0



μ $verschiebt\ auf\ X-Achse$ / σ^2 $ändert\ die\ Steilheit\ des\ Gipfels$

Fläche unter der Kurve muss immer gleich 1

Kumulative Verteilfunktion CDF:

$$\phi_{\mu,\sigma}(x)=P(X\le x)=\int_{-\infty}^{x}\varphi_{\mu,\sigma}(t)dt$$
$$=\frac{1}{\sigma\cdot\sqrt{2\cdot\pi}}\cdot\int_{-\infty}^{x}e^{-\frac{1}{2}\frac{(t-\mu)^2}{\sigma^2}}dt$$

Standardnormalverteilung (Tabelle Papula S. 514)

standardisiert Zufallsvariablen X betrachtet man die Zufallsvariable:

$$U=\frac{X-\mu}{\sigma}$$

Beispiel mit standardisieren

$N(\mu,\sigma^2)$ $N(3,2)$ $\mu=3$ $\sigma^2=2$ $\sigma=\sqrt{2}$

$$P(X\le 4.84)=P\left(\frac{X-\mu}{\sigma}\le\frac{4.84-\mu}{\sigma}\right)$$
$$=P\left(U\le\frac{4.84-\mu}{\sigma}\right)=\phi(0.92)$$
$$=0.8212\ (Formelsammlung)$$

$$\phi(u)=Standardnormalverteilung$$

Für negative Argumente verwendet man die Formel:

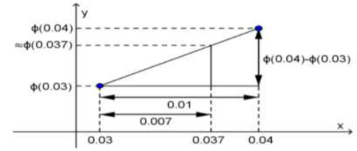
$\phi(-u)=1-\phi(u)$ $(u>0)$

$Für\ u\ge 4\ ist\ \phi(u)\cong 1$

Linke Spalte = erste Nachkommastelle;

Horizontale Spalte = zweite Nachkommastelle

Hat die vorgegebene Grenze mehr als zwei Nachkommastellen, wird linear interpolier:



Beispiele Standardnormalverteilung

- $P(X \le 0.45) \rightarrow Papula\ Formelsammlung\ nachschlagen$
- $P(X \le -0.21) = 1 - P(X \le 0.21) = 1 - (Wert\ im\ Formelbuch)$
- $P(X \le 0.037)$: (Interpolation)
  $Steigung = \frac{0.004}{0.01} =$
  $\frac{Differenz\ zwischen\ oberer\ und\ unterer\ grenze\ (aus\ Papula\ Tabelle)}{Schrittlänge} = 0.4$
  $P(X \le 0.03) = 0.5120\ (aus\ Formelsammlung)$
  $P(X \le 0.037) = 0.5120 + (0.4 \cdot 0.007) = 0.5148$
- $P(|X| \le 1.69) = P(-1.69 \le X \le 1.69) = 2 \cdot P(U \le 1.69) - 1 \cong 2 \cdot 0.9545 - 1 = 0.9090$

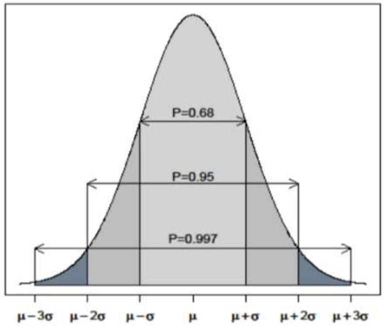
Symmetrische Wahrscheinlichkeit

Wahrscheinlichkeit ist symmetrisch zu μ und $e > 0$ ist eine Variable

$$P(\mu - e \le X \le \mu + e) = P\left(-\frac{e}{\sigma} \le U \le \frac{e}{\sigma}\right) = \phi\left(\frac{e}{\sigma}\right) - \left(1 - \phi\left(\frac{e}{\sigma}\right)\right)$$
$$= 2 \cdot \phi\left(\frac{e}{\sigma}\right) - 1$$

$P\left(-\frac{e}{\sigma} \le U \le \frac{e}{\sigma}\right)$ → Standartisation

Im Bereich von $\pm 2\sigma$ um den Mittelpunkt befinden sich 95.44% der Daten.
Bei der Normalverteilung fallen fast alle Werte innerhalb drei Standardabweichungen vom Mittelwert. Diese Regel gilt für alle Normalverteilungen (unabhängig von μ oder σ)



Beispiel der Symmetrischen Wahrscheinlichkeit

$$P(\mu - 2\sigma \le X \le \mu + 2\sigma) = 2 \cdot \phi\left(\frac{2\sigma}{\sigma}\right) - 1$$
$$= 2 \cdot \phi(2) - 1 = 2 \cdot 0.9772 - 1$$
$$= 0.9544$$

Zusammenfassung

Verteilung	W'keitsfunktion	$E[X]$ und $Var(X)$	Anwendung
Uniform $X \sim$ Unif$([a,b])$	$f(x) = \begin{cases} 0 & a < 0 \\ \frac{1}{b-a} & (a \le x \le b) \\ 0 & b > 0 \end{cases}$ R: dunif(x, min = a, max = b)	$(a+b)/2$ $(b-a)^2/12$	Wenn alle Werte in einem Bereich gleich wahrscheinlich sind.
Exponential $X \sim \text{Exp}(\lambda)$	$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \ge 0 \\ 0 & x < 0 \end{cases}$ R: dexp(x, rate = λ)	$1/\lambda$ $1/\lambda^2$	Dauer von zufälligen Zeitintervallen
Normal $X \sim \mathcal{N}(\mu, \sigma^2)$	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right)$ R: dnorm(x, mean = μ, sd = σ)	μ σ^2	Abweichungen von (Mess-) Werten, Näherung bei grossen Stichproben
Lognormal $X \sim \log\mathcal{N}(\mu, \sigma^2)$	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{(\log(x)-\mu)^2}{\sigma^2}\right)$ R: dlnorm(x, meanlog = μ, sdlog = σ)	$e^{(\mu+\frac{\sigma^2}{2})}$ $e^{(2\mu+\sigma^2)}(e^{\sigma^2}-1)$	rechtschiefe Messdaten mit nur positiven Werten

Summe von Zufallsvariablen

Summen X und Y: **Beispiele:** Augenzahl von zwei Würfeln, Unfälle in 2 Stadtteilen

Varianzen addieren ist ok, Standardabweichungen addieren ist nicht ok.
Ausnahmen für die Summenbildung:

- $X, Y\ normalverteilt:\quad S \sim N(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)$
- $X, Y\ poissonverteilt:\quad S \sim Pois(\lambda_x + \lambda_y)$
- $X, Y\ binomialverteilt\ mit\ gleichem\ p:\quad S \sim Bin(n_x + n_y, p)$

Seien a und b skalare grössen und S=X+Y eine Zufallsvariable, dann gilt immer:

Erwartungswert:

$E(S) = E(X) + E(Y) \quad E(aX + b) = aE(X) + b$

Varianz:

$V(S) = V(X) + V(Y) + 2 \cdot Cov(X,Y)$
$V(aX + b) = a^2 V(X)$

$Cov(X,Y)$ beschreibt linearen Zusammenhang der beiden Zufallsvariablen

Zentraler Grenzwertsatz

$Mittelwert\ von\ n >$
$2\ Zufallsvariablen\ mit\ identischer\ Verteilung$

$$\bar{X} \sim N_n\left(\mu\, ; \frac{\sigma^2}{n}\right) \qquad für\ n \rightarrow \infty$$

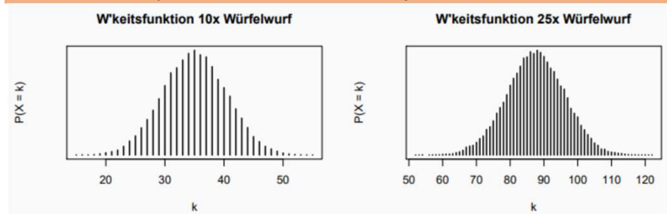
$\mu = Erwartungswert = E(X)$
$\frac{\sigma^2}{n} = Varianz = V(X)$

Je grösser das n umso kleiner die Streuung der Daten.

$$\bar{X} \sim N\left(\mu_X\, ; \frac{\sigma_X^2}{n}\right) mit\ \bar{\bar{X}} = \frac{1}{n}\sum_{i=1}^{n} X_i = S_n$$

Summen von $n > 2$ unabhängige Zufallsvariablen mit identischer Verteilung

$$S_n \sim N(n \cdot \mu, n \cdot \sigma^2) \quad für\ n \rightarrow \infty$$



Binomial- als auch die Poissonverteilung durch eine Normalverteilung annäherbar.

Approximation durch die Normalverteilung: Binomialverteilung

$X = X_1 + X_2 + X_3 + \cdots + X_n$ = Summe von n Bernoulli Experimenten

$\mu = n \cdot p$ und $\boldsymbol{\sigma^2 = n \cdot p \cdot q} = n \cdot p \cdot (1-p)$

$$P(a \le X \le b) = \sum_{x=a}^{b} P(X=x)$$

$$\cong \phi_{\mu,\sigma^2}\left(b+\frac{1}{2}\right) - \phi_{\mu,\sigma^2}\left(a-\frac{1}{2}\right)$$

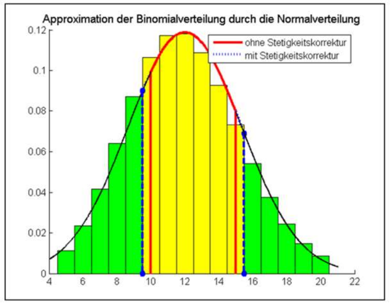
$$P(a < X < b) \cong \phi_{\mu,\sigma^2}\left(b-\frac{1}{2}\right) - \phi_{\mu,\sigma^2}\left(a+\frac{1}{2}\right)$$

Bemerkung: man betrachtet die Fläche zwischen $a - \frac{1}{2}$ und $b + \frac{1}{2}$

Faustregel: Approximation gilt, wenn $\boldsymbol{n \cdot p \cdot q > 9}$

Beispiel Binominalverteilung durch die Normalverteilung

$$P(10 \le X \le 15) = \sum_{x=a}^{b} P(X=x) \approx \phi_{12,\sqrt{11.28}}(15.5) - \phi_{12,\sqrt{11.28}}(9.5) \approx 0.6230$$



Approximation durch die Normalverteilung: Poissonverteilung

X sei poissonverteilt mit Parameter λ.

$\mu = \lambda$ und $\sigma^2 = \lambda$

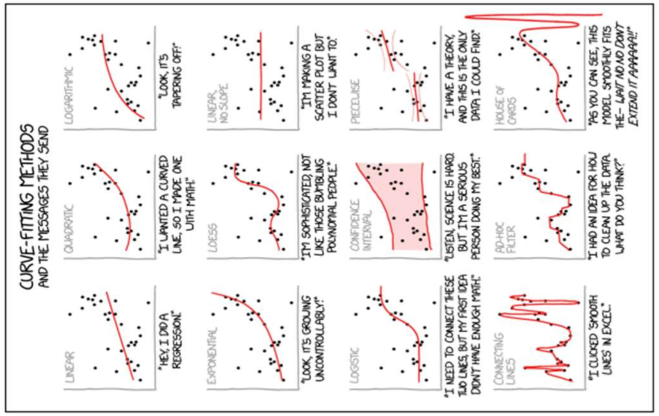
$$P(a \le X \le b) = \sum_{x=a}^{b} P(X=x)$$

$$\cong \phi_{\mu,\sigma^2}\left(b+\frac{1}{2}\right) - \phi_{\mu,\sigma^2}\left(a-\frac{1}{2}\right)$$

Faustregel: Approximation gilt, wenn $\lambda > 9$

6. Die Methode der kleinsten Quadrate

6.1 Einführung



Alltägliche Frage: Zusammenhang metrische Zielgrösse spezieller Interesse und mehreren Einflussgrössen?

Beispiele: Treibstoffverbrauch abhängig von Route, Wind,…; Wohnungsmiete abhängig von Grösse, Ort,…; Einkommen abhängig von Ausbildung, Alter,…;

Regression: → Quantitative Beschreibung dieser Zusammenhänge

→Am weitesten verbreitete statistische Methode

Mathematisches Beispiel Flugpassagiere

Flughafen Zürich werden jeden Monat Passagier-aufkommen und Anzahl Flugbewegungen erhoben.

Zielgrösse: $y_i = Anzzahl\ Passagiere\ im\ Monat\ i$

Erklärende Variablen für Monat i:

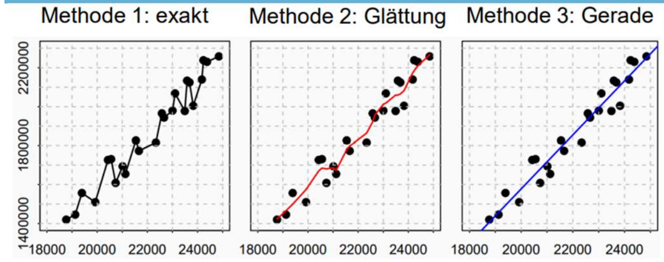
$x_i^{(1)} = Anzahl\ Flugzeugbewegung;\ x_i^{(2)} = Anzahl\ Cargoflüge;\ x_i^{(3)} = Jahr$

Gesucht wird Funktion *h(.)*:

$$y_i = h(x_i^{(1)}, x_i^{(2)}, x_i^{(3)})$$

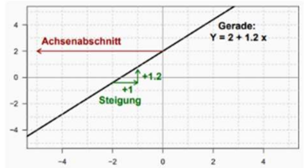
y_i *für jede Beobachtung i beschreiben lässt.*

6.2 Lineare Regression



Die Kurve nicht das Rauschen, sondern den «wirklichen Zusammenhang» beschreiben.

- Punkte verbinden schlechtes Modell (erstes Bild)
- Eine Gerade einfachste Modell (drittes Bild)



$y = h(x) = d + m \cdot x = Abschnitt + Steigung \cdot x$

$\boldsymbol{d = \bar{y} - m \cdot \bar{x}}$

$y = Anzahl\ PAX$

$x = Anzahl\ Flugbewegungen$

$$y_i = d + m \cdot x_i + \varepsilon_i \qquad für\ alle\ i = 1, \dots, n$$

$y_i = Zielvariable\ der\ i-ten\ Beobachtung$

$x_i = erklärende\ Grösse\ der\ i-ten\ Beobachtung$

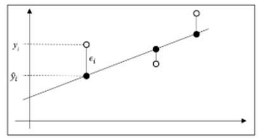
$d, m = sind\ die\ Regressionskoeffizienten\ welche\ bestimmt\ werden\ müssen$

$\varepsilon_i = zufäll.\ Rest\ oder\ Fehler\ (Abweichung\ zw.\ Beobachtung\ u.\ Geraden\ o.\ Zufallsgrösse)$

Anpassung der Geraden

$$\varepsilon_i = y_i - \hat{y}_i \quad \rightarrow$$

$Residuen = Gemessenes\ y - Geschätztes\ y\ auf\ Geraden$



Methode der Kleinsten Quadrate (KQM): Regressionsgerade angepasst, dass Summe der quadrierten Abweichungen ($\boldsymbol{\varepsilon_i}$) minimal wird

$$Q(d,m) = \sum_{i=1}^{n} (e_i)^2 = \sum_{i=1}^{n} \left(y_i - (d + m \cdot x_i)\right)^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = minimal$$

$\hat{y}_i = d + m \cdot x_i \qquad \varepsilon_i = y_i - \hat{y}_i$

$Q(d,m)$ gibt an, wie gut die Regressionsgerade zu den Daten passt (je kleiner Q, umso besser)

Minimierung der Varianz der Residuen s_ε^2

$$m = \frac{\sum_{i=1}^{n}(y_i - \bar{y})\cdot(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{s_{xy}}{s_x^2} = \frac{Kovarianz}{Varianz}$$

$$s_\varepsilon^2 = \frac{1}{n}\cdot\sum_{i=1}^{n}(y_i - m\cdot x_i - d)^2$$

$\bar{x} = Datenschwerpunkt\ von\ x$

$\bar{y} = Datenschwerpunkt\ von\ y$

$s_{xy} = Kovarianz\ von\ x$

$s_x^2 = Varianz\ von\ x$

- Das Verfahren ist einfach, da die Lösung explizit als Funktion der Datenpunkte geschrieben werden kann
- Die Summe der Abweichungen addiert sich zu null
- Die Gerade geht durch den Daten-Schwerpunkt $(\bar{x}, \bar{y})$

Varianz der x_i-Werte

$$s_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \left(\frac{1}{n}\sum_{i=1}^{n}x_i^2\right) - \bar{x}^2$$

Varianz. der y_i-Werte

$$s_y^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 = \left(\frac{1}{n}\sum_{i=1}^{n}y_i^2\right) - \bar{y}^2$$

Kovarianz

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

arithmetische Mittelwerte

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n}x_i \text{ und } \bar{y} = \frac{1}{n}\sum_{i=1}^{n}y_i$$

Verteilung der Fehler

$y_i = d + m\cdot x_i + \varepsilon_i$

Annahme: $\varepsilon_i \sim N(0, s_\varepsilon^2)$

$s_\varepsilon^2 = minimale\ quadratische\ Varianz$

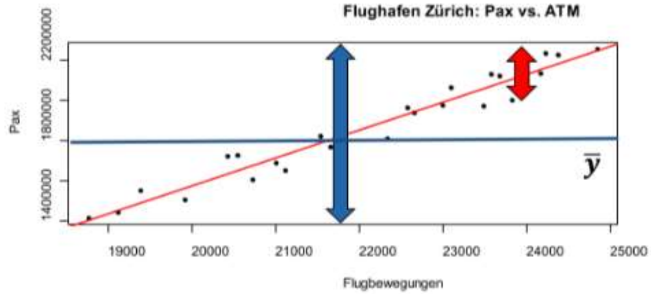
Varianz bleibt immer gleich gross

Wenn innerhalb der Punktewolke:

$E(\varepsilon_i) = 0 \qquad V(\varepsilon_i) = s_\varepsilon^2$

Bestimmtheitsmass R^2

$$R^2 = \frac{s_{\hat{y}}^2}{s_y^2} = \frac{Erklärte\ Varianz}{totale\ Varianz} = \frac{s_{xy}^2}{s_x^2\cdot s_y^2} = r_{xy}^2$$



Je kleiner der rote Pfeil im Vergleich zum blauen ist, desto grösser ist der Erklärungsgehalt der Regressionsgerade

$R^2 = 0.95$ Vernünftiger Wert für R: $R^2 > 0.4$

Zusammenfassung

x	$\bar{x} = \frac{1}{n}\cdot\sum()$
y	$\bar{y} = \frac{1}{n}\cdot\sum()$
$(x - \bar{x})$	$s_x^2 = \frac{1}{n}\sum()^2$
$(y - \bar{y})$	$s_y^2 = \frac{1}{n}\sum()^2$
$(x - \bar{x})\cdot(y - \bar{y})$	$s_{xy} = \sum()$

Totale Varianz $s_y^2 = \frac{1}{n}\sum(y - \bar{y})^2 = \frac{1}{n}\cdot\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \frac{1}{n}\cdot\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 = s_\varepsilon^2 + s_{\hat{y}}^2$

Erklärte Varianz $s_{\hat{y}}^2 = \frac{1}{n}\cdot\frac{s_{xy}^2}{s_x^2}$

Residuen Varianz $s_\varepsilon^2 = s_y^2 - s_{\hat{y}}^2$

Korrigierte Varianz $s_{x,korr}^2 = \frac{1}{n-1}\cdot\sum_{i=1}^{n}(x_i - \bar{x})^2$

Bestimmungsmass $R^2 = \frac{s_{\hat{y}}^2}{s_y^2} = \frac{\frac{s_{xy}^2}{s_x^2}}{s_y^2}$

Pearson Korr. $r = \sqrt{R^2}\ aufpassen\ mit + und - !$

Gerade bestimmen $y(x) = m\cdot x + d$

$m = \frac{s_{xy}}{s_x^2} \qquad d = \bar{y} - m\cdot\bar{x} \quad oder \quad d = m\cdot\bar{x} - \bar{y}$

$x(y) = m\cdot y + d \qquad m = \frac{s_{xy}}{s_y^2}$

$d = \bar{x} - m\cdot\bar{y} \quad oder \quad d = m\cdot\bar{y} - \bar{x}$

6.3 Nichtlineares Verhalten

Ausgangsfunktion (nicht linear)	Transformation (linearisiert)
$y = q\cdot x^m$	$\ln(y) = \ln(q) + m\cdot\ln(x)$
$y = q\cdot m^x$	$\ln(y) = \ln(q) + \ln(m)\cdot x$
$y = q\cdot e^{m\cdot x}$	$\ln(y) = \ln(q) + m\cdot x$
$y = e^{q\cdot\frac{m}{x}}$	$\ln(y) = q - m\cdot U \;\rightarrow\; U = \frac{1}{x}$
$y = \frac{1}{q + m\cdot x}$	$\frac{1}{y} = V = q + m\cdot x$
$y = q + m\cdot\ln(x)$	$y = q + m\cdot U \;\rightarrow\; U = \ln(x)$
$y = \frac{1}{q\cdot m^x}$	$\ln\left(\frac{1}{y}\right) = \ln(q) + \ln(m)\cdot x$
Rücktransformation	$d = \ln(q) \;\rightarrow\; q = e^d$

(Am Schluss Rücktransformation nicht vergessen!)

6.4 Allgemeines Vorgehen

Logarithmierens:

Geldbeträge, Konzentrationen (Chemie), physikalische oder chemische Grössen (Daten müssen strikt positiv sein, ***0 ist ein Problem***)

Wurzeltransformationen

Bei Poisson-Verteilung; Varianz = Mittelwert = λ → im Regressionskontext inhomogene Varianzen

Bei nicht linearen Regressionen, Variablen transformieren ($x \rightarrow \tilde{x}$ und $y \rightarrow \tilde{y}$), mit transformierten Variablen die lineare Regression durchspielen

7 Schliessende Statistik – Parameter- und Intervallschätzung

7.1 Zufallsstichproben

Ein Schätzer ist eine Funktion der Zufallsstichprobe

$$\Theta = g(X_1, \ldots, X_n)$$

Setzt man einen konkrete Stichprobe ein, so ist

$$\hat{\Theta} = g(x_1, \ldots, X_n)$$

ein Schätzwert für den Parameter Θ

Θ ist also selber eine Zufallsvariable, die einer gewissen Verteilung folgt.

7.2 Parameterschätzungen

Beispiele von Parameter

Fragestellung	Wie gross ist der Anteil Ja-Stimmen?
Merkmal	Ja oder Nein
Typ der Verteilung	Bernoulli-Verteilung

Fragestellung	Wie genau wird der Sollwert eingehalten?
Merkmal	Abweichung vom Sollwert
Typ der Verteilung	Normalverteilung

Einige Schätzfunktionen:

- Schätzfunktionen für μ: arithmetisches Mittel, Median
- Schätzfunktionen für σ: empirische Standardabweichung
- Schätzfunktionen für p: relative Häufigkeit

Kriterien für eine optimale Schätzfunktion

Schätzfunktion Θ eines Parameters θ heisst *erwartungstreu*, wenn gilt:

$$E(\Theta) = \theta$$

zwei erwartungstreue Schätzfunktionen Θ_1 und Θ_2 desselben Parameters θ. Θ_1 *effizienter* als Θ_2, falls gilt:

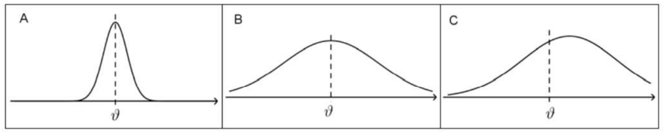
$$V(\theta_1) < V(\theta_2)$$

Eine Schätzfunktion Θ heisst *konsistent*, wenn gilt:

$$E(\Theta) \to \theta \text{ und } V(\Theta) \to \theta \text{ für } n \to \infty$$

Beispiele zu erwartungstreu, effizient und konsistent

1. Dichtefunktion einer nicht erwartungstreuen Schätzfunktion.
2. Dichtefunktion einer erwartungstreuen und effizienten Schätzfunktion.
3. Dichtefunktion einer erwartungstreuen, aber weniger effizienten Schätzfunktion. θ



Schätzer – Erwartungswert

Schätzfunktion $\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i$

Schätzwert $\hat{\mu} = \bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$

$E(\bar{X}) = E(X_i) \quad \to \quad \bar{X}$ *ist erwartungstreu*

$V(\bar{X}) = \frac{V(X_i)}{n}$ für $\xrightarrow{n\to\infty} 0 \quad \to \bar{X}$ *ist konsistent*

Schätzer – Varianz

Schätzfunktion $S^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (X_i - \bar{X})^2$

Schätzwert $\hat{\sigma}^2 = s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2$

$E(S^2) = V(X_i) \quad \to \quad S^2$ *ist erwartungstreu*

$V(S^2) = \frac{E((X_i-\mu)^4)}{n} - \frac{(n-3)\cdot V(X_i)^2}{n\cdot(n-1)}$ für $\xrightarrow{n\to\infty} 0$

Spezialfall: Anteilswert p einer Bernoulli-Verteilung

Schätzfunktion $\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i$

Schätzwert $\hat{p} = \bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$

$E(\bar{X}) = E(X_i) = p \quad \to \bar{X}$ *ist erwartungstreu*

$V(\bar{X}) = \frac{V(X_i)}{n} = \frac{p\cdot(1-p)}{n}$ für $\xrightarrow{n\to\infty} 0 \quad \bar{X}$ *ist konsistent*

Schätzfunktionen

Verteilung	Kenngrössen	Möglicher Schätzer für die Parameter
Binomial: $X \sim Bin(n,p)$	$E(X) = n\cdot p$ $V(X) = n \cdot p \cdot (1-p)$	$\hat{p} = \bar{x}$
Poisson: $X \sim Pois(\lambda)$	$E(X) = \lambda$ $V(X) = \lambda$	$\hat{\lambda} = \bar{x}$
Uniform: $X \sim Unif(a,b)$	$E(X) = \frac{a+b}{2}$ $V(X) = \frac{(b-a)^2}{12}$	$\hat{a} = \bar{x} - sd(x) \cdot \sqrt{3}$ $\hat{b} = \bar{x} + sd(x) \cdot \sqrt{3}$ $sd(x) = korr.Std.abw$
Normal: $X \sim N(\mu, \sigma^2)$	$E(X) = \mu$ $V(X) = \sigma^2$	$\hat{\mu} = \bar{x}$ $\hat{\sigma} = sd(x)$

7.3 Vertrauensintervalle

Graphische Methoden zur Überprüfung von Verteilungsannahmen

Idee: Wir vergleichen die Quantile von

- **Theoretische** kumulative Wahrscheinlichkeitsfunkt
$$F(x) = P(X \le x)$$
- **Empirische** kumulative Wahrscheinlichkeitsfunktion
$$\hat{F}_n(x) = Anzahl\ Beobachtungen\ /\ n$$

Konfidenzintervall (KI)

Transformation von $\bar{X}_n$ ergibt eine neue Zufallsvariable Z_n , die standardnormalverteilt ist $(N(1{,}0))$

$$P\left(q_{\frac{\alpha}{2}} \le Z_n \le q_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

$$\alpha = 1 - p \to (1 - gegebene\ \%\ Angabe)$$

$$Z_n = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \qquad \sim N(0{,}1)$$

Konfidenzintervall für μ einer Normalverteilung mit σ^2 bekannt

Untere Grenze: $\bar{X}_n - \frac{\sigma}{\sqrt{n}} \cdot q_{1-\frac{\alpha}{2}}$

Obere Grenze $\bar{X}_n + \frac{\sigma}{\sqrt{n}} \cdot q_{1-\frac{\alpha}{2}}$

$q_{1-\frac{\alpha}{2}}$ wird aus der Tabelle für Normalverteilung bestimm: Quantile Standardnormalver *(Papula S.516)*

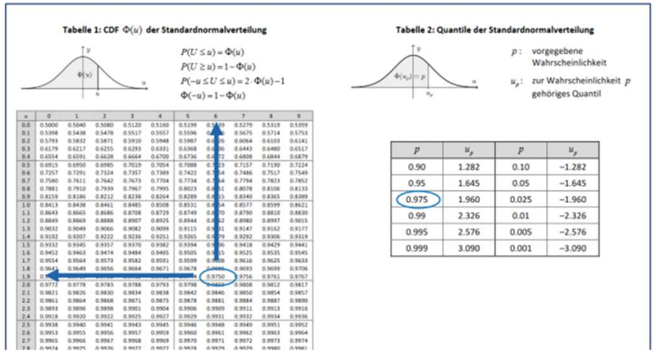


Tabelle 1: CDF Φ(u) der Standardnormalverteilung

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

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	[illegible]	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	[illegible]	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	[illegible]	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	[illegible]	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	[illegible]	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	[illegible]	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	[illegible]	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	[illegible]	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	[illegible]	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	[illegible]	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	[illegible]	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	[illegible]	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	[illegible]	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	[illegible]	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	[illegible]	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	[illegible]	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	[illegible]	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	[illegible]	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	[illegible]	0.9693	0.9699	0.9706
1.9	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	[illegible]	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

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

$p = Anzahl\ Quantil, \qquad u_p = Wert\ des\ Quantils$

$(1-\alpha) - Konfidenzintervall =$ $(Untere\ Grenze\ \ ;\ \ Obere\ Grenze)$ stets symmetrisch um $\bar{X}_n$

Konfidenzintervall für μ einer Normalverteilung mit σ^2 unbekannt

Varianz mittels Stichprobenvarianz schätzten:

$$\hat{\sigma}^2 = s_n^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2$$

Standardisierter Mittelwert:

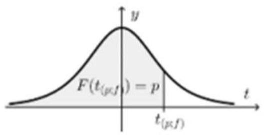
$T_n = \sqrt{n} \cdot \frac{\bar{X}_n - \mu}{s_n}$ $\qquad$ T_n war vorhin das Z_n

Untere Grenze: $\qquad \bar{X}_n - \frac{s_n}{\sqrt{n}} \cdot qt_{\left(1-\frac{\alpha}{2};(n-1)\right)}$

Obere Grenze: $\qquad \bar{X}_n - \frac{s_n}{\sqrt{n}} \cdot qt_{\left(1-\frac{\alpha}{2};(n-1)\right)}$

$qt_{\left(1-\frac{\alpha}{2};(n-1)\right)}$ wird aus der Tabelle Quantile der **t-Verteilung** nach «Student» *(Papula S.520)*bestimmt

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



p: vorgegebene Wahrscheinlichkeit

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

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

f	p = 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	[illegible]	[illegible]	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

$(1-\alpha) - Konfidenzintervall =$
$(Untere\ Grenze \ ;\ Obere\ Grenze)$

Vertrauensintervalle, Beispiel: Kundenzufriedenheit einer Airline

Unter 112 zufällig ausgewählten Kunden sind 26 unzufrieden. Es soll der Anteil p der unzufriedenen Personen unter allen Kunden geschätzt werden. Gesucht ist auch ein 95% Vertrauensintervall für p .

Zufallsvariable: $\qquad X_{unzufrieden}: X \sim Bin(n,p)$

Schätzung für den Anteil

$$\hat{p} = \frac{Anzahl\ Unzufriedene}{n} = \frac{26}{112} = 0.232$$

Für das 95% Konfidenzintervall wird eine obere und untere Grenze gesucht, sodass:

$P(p_u \le p \le p_o) = 1 - \alpha = 0.95$

Übersicht Vertrauensintervalle

$\gamma = 1 - \alpha, \qquad f = Freiheitsgrade, \qquad c = q_{1-\frac{\alpha}{2}},$

$e = hilfsfunktion = alles\ ausser\ \bar{X}\ \left(e.g. = c \cdot \frac{\sigma}{\sqrt{n}}\right)$

	(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	$\hat{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ X_i 0/1-wertig mit $P(X_i = 1) = p$	$U = \frac{\hat{X} - p}{\sqrt{p(1-p)/n}}$	**Standardnormalverteilung** (näherungsweise), Tabelle 2 $c = u_q$ mit $q = \frac{1+\gamma}{2}$	$\Theta_u = \hat{X} - c \cdot \sqrt{\frac{\hat{X} \cdot (1-\hat{X})}{n}}$ $\Theta_o = \hat{X} + c \cdot \sqrt{\frac{\hat{X} \cdot (1-\hat{X})}{n}}$
5	beliebig mit $n > 30$	μ, σ^2	wie im Fall 1 (gegebenenfalls mit s als Schätzwert für σ) bzw. im Fall 3			

Graphen und Diagramme