

Lineare Regression Wahrscheinlichkeitsverteilung

Peter Büchel

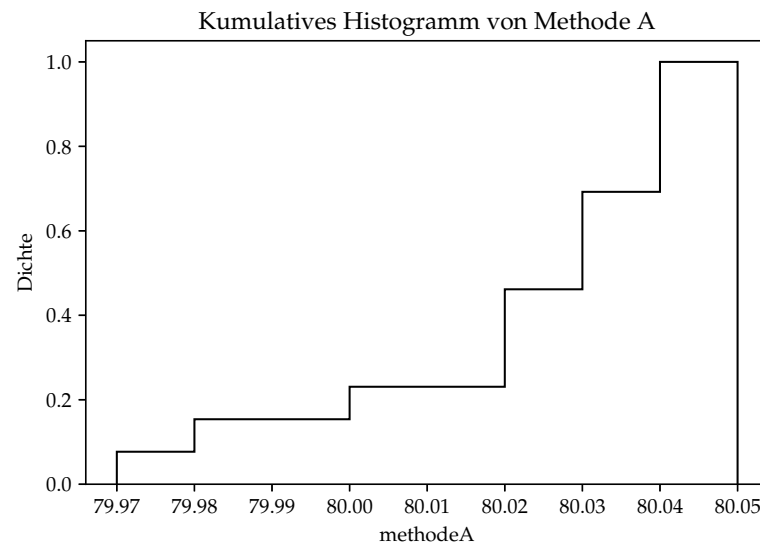
HSLU I

Stat: SW03

Empirische kumulative Verteilungsfunktion

- Empirische kumulative Verteilungsfunktion $F_n(\cdot)$ ist eine Treppenfunktion, mit:
 - ▶ Links von $x_{(1)}$ ist die Funktion gleich null
 - ▶ Bei jedem $x_{(i)}$ wird ein Sprung der Höhe $\frac{1}{n}$ gemacht
 - ▶ Wert kommt mehrmals vor \rightarrow Sprung entsprechendes Vielfache von $\frac{1}{n}$

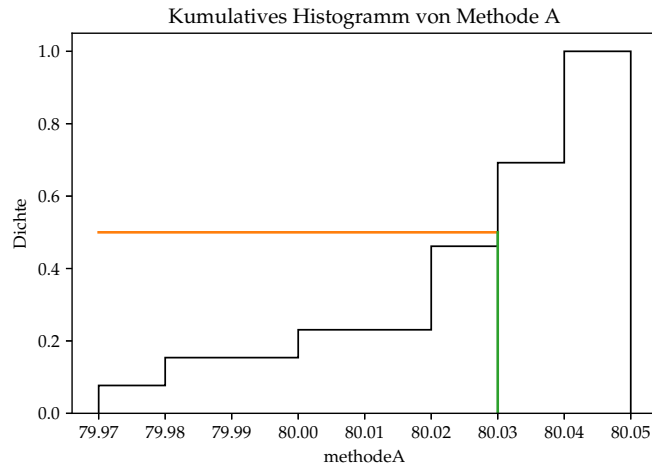
- Beispiel: Kumulative Verteilungsfunktion der Methode A



- Abbildung entsteht wie folgt:
 - ▶ Jeder Beobachtung wird ein Dichtewerte von $\frac{1}{13}$ zugeordnet
 - ▶ Links von 79.97 ist Funktion 0 (es hat keinen kleineren Beobachtungswert)
 - ▶ Bei 79.97 macht die Funktion einen Sprung auf $n = \frac{1}{13} \approx 0.077$
 - ▶ Funktion bleibt dann gleich bis 80.00, da es vorher keinen zusätzlichen Beobachtungswert gibt
 - ▶ Bei 80.00 macht die Funktion wieder einen Sprung um 0.077 nach oben, weil es dort einen Messwert hat
 - ▶ Bei 80.02 macht die Funktion einen Sprung um $3 \cdot 0.077$ nach oben, da es dort 3 Beobachtungswerte gibt
 - ▶ usw.
 - ▶ Bei 80.05 letzten Sprung \rightarrow Funktionswert wird 1

- Was kann man aus der kumulativen Verteilungsfunktion herauslesen?

- Bei 0.5 auf vertikaler Achse werden gerade die Hälfte aller Werte aufsummiert



- Zeichnen von 0.5 horizontale Linie → grüne Linie in Abbildung schneidet kumulative Verteilungsfunktion bei 80.03
- Das entspricht gerade dem Median
- Dort, wo die kumulative Funktion steil, viele Beobachtungswerte
- D.h.: die meisten Beobachtungswerte liegen hier zwischen 80.02 und 80.04
- Die Werte entsprechen aber gerade dem unteren und oberen Quartil

Python

```
methodeA.plot(kind="hist", cumulative=True, histtype="step",
normed=True, bins=8, edgecolor="black")
```

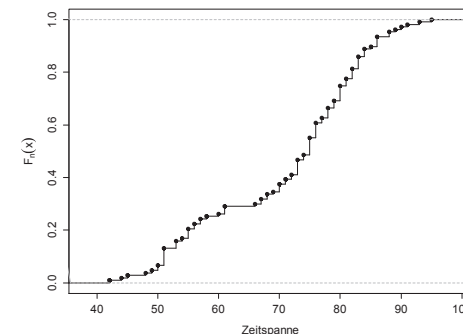
- Kumulative Verteilungsfunktion: `cumulative=True`

Empirische kumulative Verteilungsfunktion

- Empirische kumulative Verteilungsfunktion ist definiert als der *Anteil der Punkte kleiner als ein bestimmter Wert*

$$F_n(x) = \frac{1}{n} \text{Anzahl}\{i \mid x_i \leq x\}$$

- Kumulative Verteilungsfkt. für Zeitspanne im Geysir-Datensatz



Sprunghöhe $1/n$ bei Beobachtungen x_i (bzw. ein Vielfaches davon, wenn es mehrere Beobachtungen mit dem gleichen Wert x_i gibt).

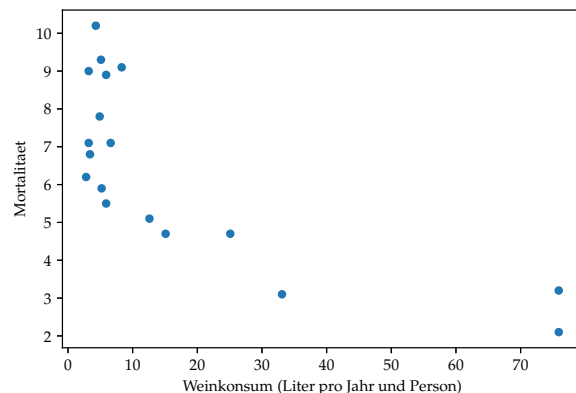
- Betrachten nun *paarweise* beobachtete Daten: *Zwei* Messgrößen pro Messeinheit

$$\begin{array}{c} x_1, \dots, x_n \\ \updownarrow \\ y_1, \dots, y_n \end{array}$$

- Weinkonsumation (Liter pro Person pro Jahr) und Mortalität aufgrund von Herz-Kreislauferkrankung (Todesfälle pro 1000) in 18 Ländern
- Eruptionsdauer (y_i) und die Zeitspanne (x_i) zum vorangehenden Ausbruch des Old Faithful Geysir

Land	Weinkonsum	Mortalität Herzerkrankung
Norwegen	2.8	6.2
Schottland	3.2	9.0
Grossbritannien	3.2	7.1
Irland	3.4	6.8
Finnland	4.3	10.2
Kanada	4.9	7.8
Vereinigte Staaten	5.1	9.3
Niederlande	5.2	5.9
New Zealand	5.9	8.9
Dänemark	5.9	5.5
Schweden	6.6	7.1
Australien	8.3	9.1
Belgien	12.6	5.1
Deutschland	15.1	4.7
Österreich	25.1	4.7
Schweiz	33.1	3.1
Italien	75.9	3.2
Frankreich	75.9	2.1

Zweidimensionales Streudiagramm



- Plot deutet an, dass hoher Weinkonsum weniger Sterblichkeit wegen Herz-Kreislauferkrankungen zur Folge hat
- Kann Zufall sein (keine Kausalität)
- Heisst *nicht*, dass Weinkonsum gesund ist (Leber!)

Streudiagramm mit Python

```
import pandas as pd
from pandas import DataFrame, Series
import numpy as np

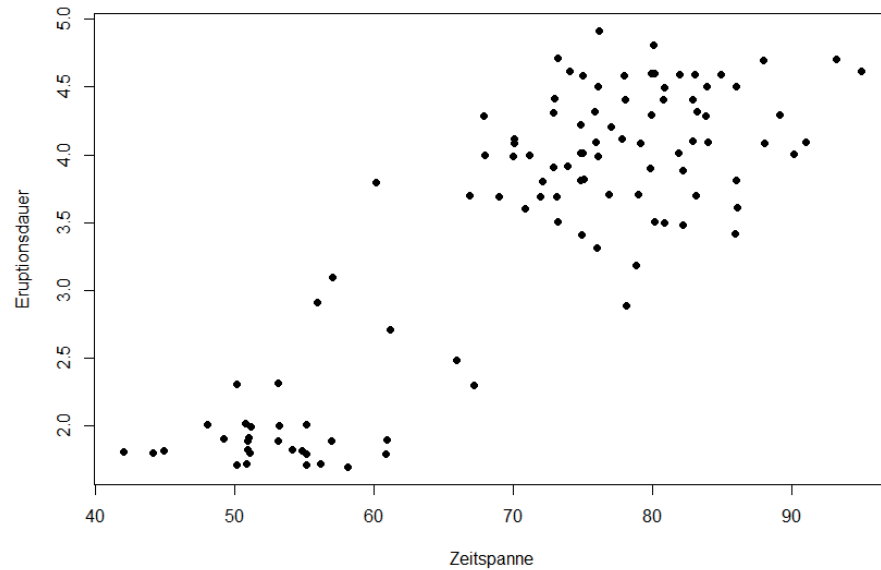
mort = DataFrame({
    "wine": ([2.8, 3.2, 3.2, 3.4, 4.3, 4.9, 5.1, 5.2, 5.9,
              5.9, 6.6, 8.3, 12.6, 15.1, 25.1, 33.1, 75.9, 75.9]),
    "mor": ([6.2, 9.0, 7.1, 6.8, 10.2, 7.8, 9.3, 5.9, 8.9,
            5.5, 7.1, 9.1, 5.1, 4.7, 4.7, 3.1, 3.2, 2.1])
})

mort.plot(kind="scatter", x="wine", y="mor")

plt.xlabel("Weinkonsum (Liter pro Jahr und Person)")
plt.ylabel("Mortalitaet")

plt.show()
```

Beispiel Old Faithful



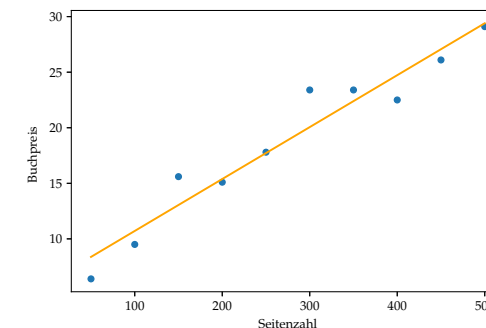
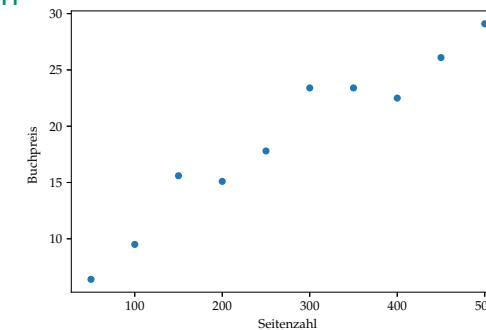
(Fiktives) Beispiel für Lineare Regression

- Kunde kauft in Buchhandlung 10 Bücher

	Seitenzahl	Buchpreis (SFr)
Buch 1	50	6.4
Buch 2	100	9.5
Buch 3	150	15.6
Buch 4	200	15.1
Buch 5	250	17.8
Buch 6	300	23.4
Buch 7	350	23.4
Buch 8	400	22.5
Buch 9	450	26.1
Buch 10	500	29.1

- *Beobachtung:*
 - ▶ Je dicker ein Roman ist, desto teurer ist er in der Regel
 - ▶ Es gibt Zusammenhang zwischen Seitenzahl x und Buchpreis y
- *Ziel:* Formelmässiger Zusammenhang zwischen Buchpreis und Seitenzahl
- Vorhersagen über Buchpreis für Bücher mit Seitenzahlen, die in Liste nicht auftauchen

Streudiagramm und Regressionsgerade



Regressionsgerade und Residuum

- Vermutung: Gerade scheint recht gut zu den Daten zu passen
- Diese Gerade hätte die Form:

$$y = a + bx$$

mit

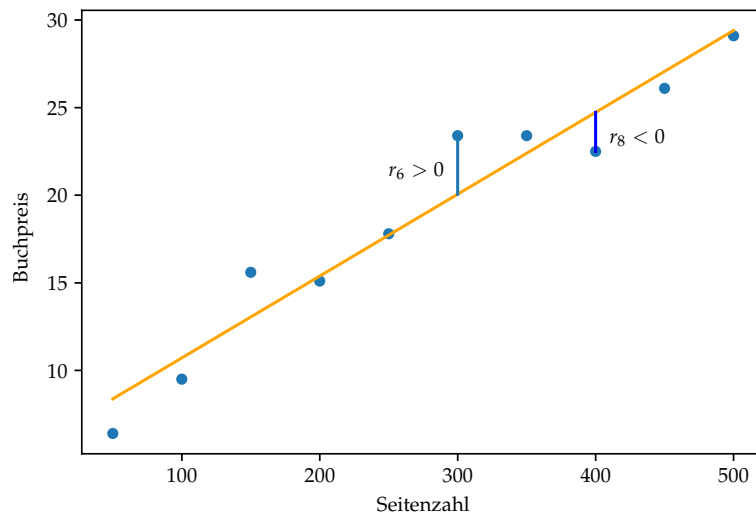
- ▶ y : Buchpreis; x : Seitenzahl
- ▶ a : Grundkosten des Verlags, b : Kosten pro Seite
- Problem: Gerade finden, die möglichst gut zu allen Punkten passt?

- Möglichkeit: Vertikale Abstände zwischen Beobachtung und Gerade zusammenzählen
- Dabei sollte eine kleine Summe der Abstände eine gute Anpassung bedeuten
- Abstände von Messpunkten zu Geraden → neuer Begriff:

Residuum

Der vertikalen Abstand zwischen einem Beobachtungspunkt (x_i, y_i) und der Geraden (der Punkt auf der Geraden ist $(x_i, a + bx_i)$) heisst *Residuum*:

$$r_i = y_i - a - bx_i$$



- Beispiel: Residuen r_6 und r_8 für *diese* Gerade in Abbildung
- Residuum r_6 positiv, da Punkt überhalb der Gerade
- Entsprechend ist $r_8 < 0$
- Gerade $y = a + bx$ so bestimmen, dass die Summe

$$r_1 + r_2 + \dots + r_n = \sum_i r_i$$

minimal wird

- Minimierung von $\sum_i r_i$ hat aber eine **gravierende Schwäche**: Falls Hälfte der Punkte weit über der Geraden, die andere Hälfte weit unter der Geraden liegen: Summe der Abstände etwa null
- Dabei passt die Gerade gar nicht gut zu den Datenpunkten!

Methode der kleinsten Quadrate

- Eine andere Möglichkeit besteht darin, die Quadrate der Abweichungen aufzusummieren, also

$$r_1^2 + r_2^2 + \dots + r_n^2 = \sum_i r_i^2$$

- Parameter a und b so wählen, dass diese Summe minimal wird
- **Python** berechnet für Beispiel die Werte $a = 6.04$ und $b = 0.047$
 - ▶ Grundkosten des Verlags sind also rund 6 SFr. (Preis des Buches für 0 Seiten)
 - ▶ Pro Seite verlangt der Verlag rund 5 Rappen
 - ▶ Geradengleichung:

$$y = 6.04 + 0.04673x$$

Bestimmung der Parameter a und b

- *Frage:* Wie berechnet der Computer die Parameter a und b ?
- Die Parameter a, b minimieren (Methode der Kleinsten-Quadrate)

$$\sum_{i=1}^n (y_i - (a + bx_i))^2$$

Die Lösung dieses Optimierungsproblem ergibt:

$$b = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$
$$a = \bar{y} - b\bar{x}$$

wobei \bar{x} und \bar{y} die Mittelwerte der jeweiligen Daten

- Diese Gerade $y = a + bx$ wird auch *Regressionsgerade* genannt

Lineare Regression mit Python

- Code:

```
b, a = np.polyfit(book["pages"], book["price"], deg=1)
print(a, b)
## 6.039999999999999 0.04672727272727273
```

- Befehl

```
np.polyfit(book["pages"], book["price"], deg=1)
```

passt ein Polynom vom Grad 1 (lineare Funktion) an Daten an

- Ausgabe von 2 Werten: der erste ist die Steigung der Geraden, der zweite der y -Achsenabschnitt
- **Python** findet also $a = 6.04$ und $b = 0.0467$

Plotten der Regressionsgerade

- Diese Gerade wird in **Python** wie folgt gezeichnet:

```
book.plot(kind="scatter", x="pages", y="price")
b, a = np.polyfit(book["pages"], book["price"], deg=1)

x = np.linspace(book["pages"].min(), book["pages"].max())

plt.plot(x, a+b*x, c="orange")

plt.xlabel("Seitenzahl")
plt.ylabel("Buchpreis")

plt.show()
```

- Der Befehl

```
x = np.linspace(book["pages"].min(), book["pages"].max())
```

erzeugt einen Vektor x der Länge 50, der als 1. Wert den Minimalwert von `pages` im Dataframe `book` hat und als letzten Wert dessen Maximalwert.

Beispiel: Buchpreis

- Mit diesem Modell: Preis für Bücher mit Seitenzahlen berechnen, die in der Tabelle nicht vorkommen
- Wieviel würde nach diesem Modell ein Buch von 375 Seiten kosten?
- $x = 375$ in die Geradengleichung oben einsetzen:

$$y = 6.04 + 0.04673 \cdot 375 \approx 23.60$$

- Das Buch dürfte also etwa CHF 23.60 kosten
- Dieses Modell ist allerdings nur begrenzt gültig
- Vor allem bei *Extrapolationen* muss man vorsichtig sein
- Möglich: Was kostet ein Buch mit einer Million Seiten?
- Oder ein Buch mit -100 Seiten? → Nicht realistisch!

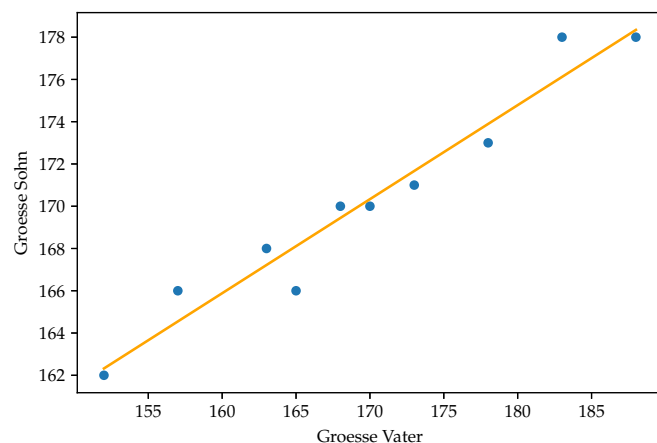
Beispiel: Körpergrösse Vater-Sohn

- Vermutung: Zusammenhang zwischen der Körpergrösse der Väter und der Grösse der Söhne
 - Der britische Statistiker Karl Pearson trug dazu um 1900 die Körpergrösse von 10 (in Wahrheit waren 1078) zufällig ausgewählten Männern gegen die Grösse ihrer Väter auf
- | | | | | | | | | | | |
|-------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Grösse des Vaters | 152 | 157 | 163 | 165 | 168 | 170 | 173 | 178 | 183 | 188 |
| Grösse des Sohnes | 162 | 166 | 168 | 166 | 170 | 170 | 171 | 173 | 178 | 178 |
- Es *scheint* einen Zusammenhang zu geben: Je grösser der Vater, desto grösser der Sohn
 - Streudiagramm: Möglicher linearer Zusammenhang besteht

- Die Punktwolke „folgt“ der Geraden

$$y = 0.445x + 94.7$$

(mit der Methode der Kleinsten Quadrate aus den Daten)



- Möglich: In Tabelle nicht vorkommende Grösse von 180 cm des Vaters, den zu erwartenden Wert für die Grösse seines Sohnes berechnen:

$$y = 0.445 \cdot 180 + 94.7 \approx 175 \text{ cm}$$

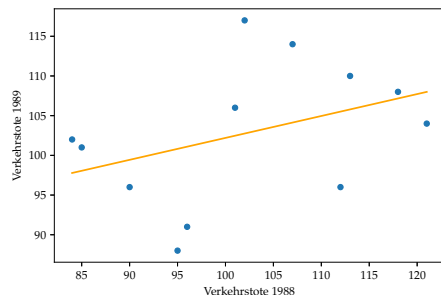
- Achtung: Formel nicht dort anwenden, wo man es *nicht* darf
- Für $x = 0$ erhält man einen Wert von 94.7
- Was heisst dies aber? Wenn der Vater 0 cm gross ist, so ist der Sohn ungefähr 95 cm gross → Macht keine Sinn!

Beispiel: Autounfälle

- Tabelle stellt einen Zusammenhang zwischen den Zahlen der Verkehrstoten her, die es 1988 und 1989 in zwölf Bezirken in den USA geben hat

Bezirk	1	2	3	4	5	6	7	8	9	10	11	12
Verkehrstote 1988	121	96	85	113	102	118	90	84	107	112	95	101
Verkehrstote 1989	104	91	101	110	117	108	96	102	114	96	88	106

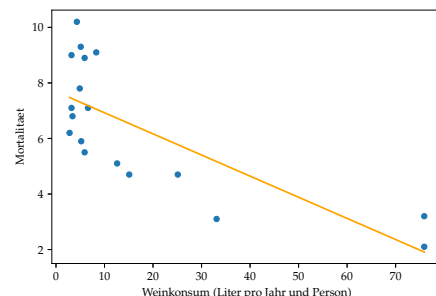
- Es besteht kein offensichtlicher Zusammenhang
- Streudiagramm: kein offensichtlicher Zusammenhang



- Zu erwarten, da es zwischen den Verkehrstoten der einzelnen Bezirke keinen Zusammenhang gibt
- In Abbildung ist noch die Regressionsgerade eingezeichnet
- Können sie zwar berechnen/einzeichnen, *aber diese macht hier gar keinen Sinn*
- Immer Berechnung und Plot vergleichen

Beispiel: Weinkonsum

- Schon gesehen: Sterblichkeit vs. Weinkonsum



- Regressionsgerade

$$y = 7.68655 - 0.07608x$$
- Zusammenhang der Daten nicht linear ist (folgt eher einer Hyperbel)
- Die Regressionsgerade sagt hier wenig über den wahren Zusammenhang aus

Wie gut passt die Regressionsgerade?

- Regressionsgerade kann (fast) immer bestimmt werden
- Letzten beiden Beispiele: Regressionsgerade sagt sehr wenig über die wirkliche Verteilung der Punkte im Streudiagramm aus
- Dafür gibt es zwei Gründe
 - Punkte folgen scheinbar gar keiner Gesetzmässigkeit
 - Punkte folgen einer nichtlinearen Gesetzmässigkeit
- Wie kann man feststellen, ob ein linearer Zusammenhang der Daten besteht oder nicht?
- Möglichkeit: Situation graphisch betrachten
- Wert angeben, der den Zusammenhang numerisch beschreibt

Empirische Korrelation

Numerischer Wert der linearen Abhängigkeit von zwei Grössen:

Empirische Korrelation

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

- Empirische Korrelation ist dimensionslose Zahl zwischen -1 und $+1$
- Misst Stärke und Richtung der *linearen Abhängigkeit* zwischen den Daten x und y
- $r = +1$: Punkte liegen auf steigender Geraden : $y = a + bx$ mit $a \in \mathbb{R}$ und ein $b > 0$
- $r = -1$: Punkte liegen auf fallender Geraden : $y = a + bx$ mit $a \in \mathbb{R}$ und ein $b < 0$
- Sind x und y unabhängig (d.h. kein Zusammenhang), so ist $r = 0$

Berechnung von Korrelation mit Python

- Seitenzahl-Preis-Beispiel mit **Python**

```
book.corr().iloc[0,1]
## 0.9681121878410434
```

- Wert sehr nahe bei 1 → starker linearer Zusammenhang
- Wert positiv → „je mehr, desto mehr“ Zusammenhang
- Der Befehl

```
book.corr()
```

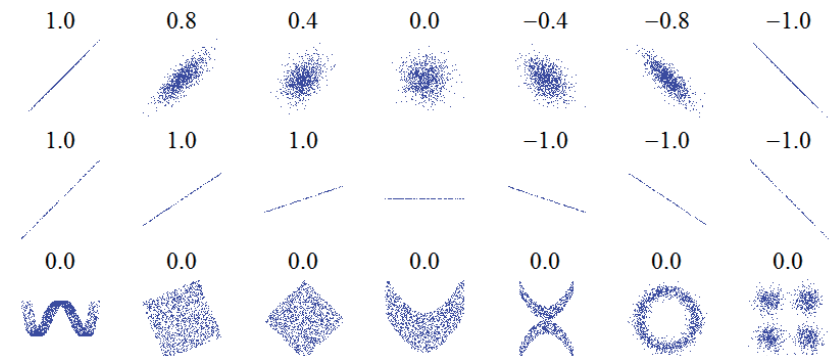
allgemeiner → Korrelationsmatrix

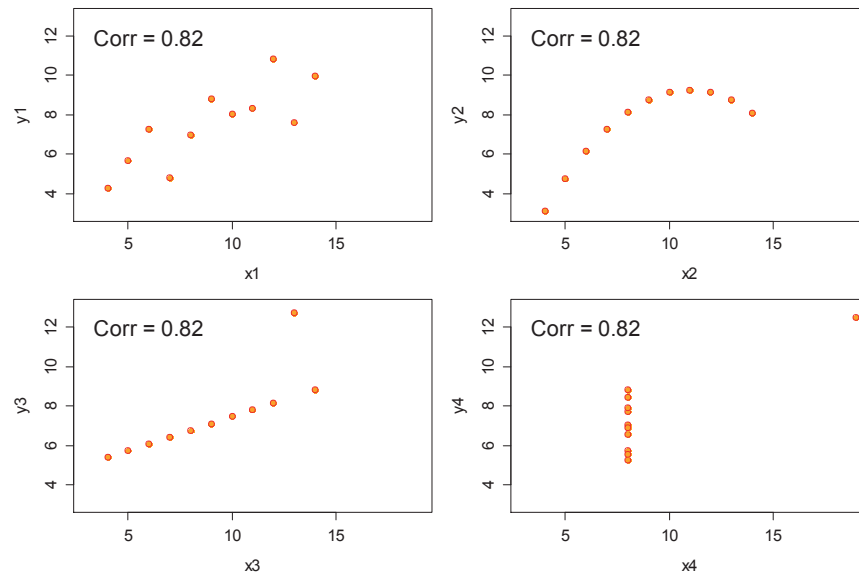
Empirische Korrelation: Beispiele

- Beispiel der Körpergrösse von Vater und Sohn: Erwarten hohen Korrelationskoeffizienten, da Daten nahe der Regressionsgerade
→ 0.973
- Verkehrsunfällen: Kein Zusammenhang → Tiefer Korrelationskoeffizienten
→ 0.386
- Weinkonsum: Keinen allzu grossen Korrelationskoeffizient (keine Gerade), aber er sollte negativ sein, da mit steigendem Weinkonsum die Mortalität sinkt
→ -0.746.

Empirische Korrelation: Bemerkungen

- Korrelation misst „nur“ den *linearen Zusammenhang*
- Man sollte daher die Daten immer auch anschauen, statt sich „blind“ auf Kennzahlen zu verlassen





- Begriff der Zufallsvariable: Spielt zentrale Rolle in der Statistik
- Beispiel: Jasskarten
 - ▶ Ein Pack Jasskarten besteht aus 36 verschiedenen Karten
 - ▶ Um beim Jassen Stiche zu vergleichen, werden den Jasskarten Zahlwerte zugewiesen
 - ▶ So hat ein König den Wert 4
 - ▶ Ohne diese Werte wären verschiedene Stiche beim Jass sehr schwierig miteinander zu vergleichen
 - ▶ Betrachten *Funktion*, die jeder Jasskarte einen Zahlwert zuordnet

▶ Also

$$\begin{array}{ll} \omega = \text{As} & \mapsto X(\omega) = 11 \\ \omega = \text{König} & \mapsto X(\omega) = 4 \\ \vdots & \vdots \\ \omega = \text{Sechs} & \mapsto X(\omega) = 0 \end{array}$$

- Dieselbe Situation kommt in der Stochastik häufig vor
- Oft wird ein Zufallsexperiment mit Zahlenwerten verknüpft
- Zu jedem Elementarereignis ω gehört ein Zahlenwert $X(\omega) = x$
- Dabei ist X eine *Funktion*, die jedem Elementarereignis ω den Zahlwert x zuordnet

- Wie in Beispiel: X ist Funktion auf dem Grundraum Ω
- Diese Funktion wird *Zufallsvariable* genannt
- Sie ordnet jedem Element des Grundraumes *eine Zahl* zu
- Vorteil: Mit den Werten der Zufallsvariable kann man rechnen
- Beispiel oben: Mit den Zahlenwerten $X(\omega)$ kann man den „Durchschnitt“ der gezogenen Karten berechnen
- Für die *Elementarereignisse* „As“, „König“ etc. macht das Wort „Durchschnitt“ keinen Sinn

Zufallsvariable

Eine *Zufallsvariable* X ist eine *Funktion*:

$$\begin{aligned} X : \quad \Omega &\rightarrow \mathbb{R} \\ \omega &\mapsto X(\omega) \end{aligned}$$

Notation X (oder auch Y, Z, \dots) ist eher ungewohnt für die Bezeichnung einer Funktion, ist aber üblich in der W'keitsrechnung

- Zufallsvariable wird mit einem *Grossbuchstaben* X (oder Y, Z) bezeichnet
- Der entsprechende *Kleinbuchstabe* x (oder y, z) stellt einen *konkreten Wert* dar, den die Zufallsvariable annehmen kann
- Für das Ereignis, bei dem die Zufallsvariable X den Wert x annimmt, schreiben wir $X = x$
- In Beispiel: Ereignis $X = 2$ entspricht „einen Under ziehen“
- Bei einer Zufallsvariable ist nicht die Funktion $X(\cdot)$ zufällig, sondern nur das Argument ω

- Je nach Ausgang des Zufallsexperiments ω erhält man einen anderen Wert $x = X(\omega)$
- x heisst dann eine *eine Realisierung* der Zufallsvariablen X
- Wird das Experiment zweimal durchgeführt erhält zweimal das gleiche Ergebnis ω , dann sind auch die realisierten Werte von X gleich
- Jasskartenbeispiel: Realisierung $X = 11$ entspricht dem Ziehen eines Asses

Diskrete Zufallsvariablen

- Hier: Zahlen, die X annehmen kann, sind *diskret*
- D.h.: Anzahl dieser Werte ist endlich (wie Jasskartenbeispiel)

$$\{0, 2, 3, 4, 10, 11\}$$

- Möglich: Unendliche Liste

$$\{2.5, 4.5, 6.5, 8.5, \dots, \}$$

- Man sagt: Zufallsvariable X ist diskret
- Insbesondere sind Anzahlen stets diskret
- Messungen meist kontinuierlich \rightarrow Mit \mathbb{R} modelliert

Wahrscheinlichkeit einer Realisierung

- Schon gesehen: W'keit $P(E)$ eines Ereignisses E berechnen
- Entsprechend: W'keit einer allgemeinen Realisierung x einer Zufallsvariable X definieren
- Beispiel: Zufallsvariable X sei der Wert einer gezogenen Jasskarte
- Wie gross die W'keit ist, dass gezogene Karte den Wert 4 hat?
- Realisierung ist in diesem Fall $X = 4$

- Bezeichnung: W'keit zur Realisierung 4

$$P(X = 4)$$

- Realisierung $X = 4$ entspricht dem Ziehen eines Königs
- D.h.: Gesucht W'keit, dass ein König gezogen wird:

$$\begin{aligned} P(X = 4) &= P(\{\omega \mid \omega = \text{ein König}\}) \\ &= P(\text{Eicheln-König}) + P(\text{Rosen-König}) + \\ &\quad + P(\text{Schellen-König}) + P(\text{Schilten-König}) \\ &= \frac{4}{36} = \frac{1}{9} \end{aligned}$$

- Vorgehen hier ausführlicher als notwendig \rightarrow verallgemeinerbar

- W'keit, dass ein König gezogen wird, ist also gleich der Summe der W'keiten die verschiedenen Könige zu ziehen
- Diese Überlegung verallgemeinern:

- ▶ Die Werte einer Zufallsvariablen X (die möglichen Realisationen von X) treten mit gewissen W'keiten auf
- ▶ Die W'keit, dass X den Wert x annimmt, berechnet sich wie folgt:

$$P(X = x) = P(\{\omega \mid X(\omega) = x\}) = \sum_{\omega; X(\omega)=x} P(\omega)$$

- Jasskartenbeispiel: $x = 4$ und ω alle möglichen Könige, deren entsprechende W'keiten aufaddiert werden

Wahrscheinlichkeitsverteilung

- Beispiel vorher: W'keit *einer* Realisierung berechnet
- Jetzt: W'keiten *aller* Realisierungen berechnen
- *Sehr wichtiger* Begriff: Wahrscheinlichkeitsverteilung

Wahrscheinlichkeitsverteilung

Für *jede* Realisierung einer Zufallsvariable die zu gehörige W'keit berechnen \rightarrow *W'keitsverteilung* dieser Zufallsvariablen

Jasskartenbeispiel

- Zufallsvariable X ist wieder der Wert einer gezogenen Jasskarte
- W'keit $P(X = 4)$ schon berechnet:

$$P(X = 4) = \frac{1}{9}$$

- W'keit $P(X = 0)$ mit der Laplace-W'keit berechnen
- Es hat unter den 36 Karten genau 16 „leere“ Karten
- Somit gilt für die W'keit $P(X = 0)$:

$$P(X = 0) = \frac{16}{36} = \frac{4}{9}$$

- Realisierung $X = 2$ entspricht dem Ziehen eines Unders
- Da es 4 von denen gibt, gilt für die W'keit $P(X = 2)$:

$$P(X = 2) = \frac{4}{36} = \frac{1}{9}$$

- W'keiten für die anderen Realisierungen analog
- *Jeder* Realisierung wird einen W'keitswert zugeordnet
- Man spricht dann von einer *Wahrscheinlichkeitsverteilung*
- W'keitsverteilung von X in Tabelle

x	0	2	3	4	10	11
$P(X = x)$	4/9	1/9	1/9	1/9	1/9	1/9

- Werte für $P(X = 1)$ oder $P(X = 178)$ sind in Tabelle *nicht* aufgeführt
- Der Grund dafür ist natürlich, dass diese Werte nicht gezogen werden können
- Ihnen wird die W'keit 0 zugeordnet

$$P(X = 1) = 0 \quad \text{oder} \quad P(X = 178) = 0$$

- Addition aller Werte der W'keitsverteilung \rightarrow muss 1 ergeben
- $$P(X = 0) + P(X = 2) + P(X = 3) + P(X = 4) + P(X = 10) + P(X = 11) = 1$$

Wahrscheinlichkeitsverteilung

- Allgemein gilt:

Die „Liste“ von $P(X = x)$ für alle möglichen Werte x heisst diskrete (*Wahrscheinlichkeits-*) *Verteilung* der diskreten Zufallsvariablen X . Dabei gilt immer

$$\sum_{\text{alle möglichen } x} P(X = x) = 1$$