Notizen MA3 WS 22/23

Noah Freising

28. November 2022

Notizen zur Vorlesung MA3 im Studiengang IB an der Hochschule Mannheim. Gehalten im Wintersemester 2022/23.

Inhaltsverzeichnis

Kombinatorik

Verteilungen

Eine Verteilungsfunktion hat den Aufbau:

$$F = P(X \leq x) \quad (1)$$
$$F : \mathbb{R} \mapsto [0;1] \quad (2)$$

F bildet also die Wahrscheinlichkeit für alle Werte der Zufallsvariablen X für $X \leq x$ ab (1). Dabei wird für jedes x eine Wahrscheinlichkeit abgebildet.

Diskrete Verteilungen

Bernoulliverteilung

Die Bernoulli-Verteilung ist eine *diskrete* Verteilung, deren Zufallsvariable X nur zwei Werte annimmt: Erfolg ($X = 1$) und Misserfolg ($X = 0$).

Für die Wahrscheinlichkeitsverteilung gilt:

$$P(X = x) = p^x (1-p)^{1-x}. \quad (3)$$

Beispiel 1: Ein Münzwurf, bei dem Kopf z.B. 1 und Zahl 0 zugeordnet wird, ist ein Beispiel für ein Bernoulli-Experiment. Bei einer fairen Münze ist $p = \frac{1}{2}$.

Binomialverteilung

EINE FOLGE VON BERNOULLI-VERSUCHEN mit der Länge n und der Trefferwahrscheinlichkeit p nennen wir *Bernoulli-Kette*. Diese ist binomialverteilt mit der Washrscheinlichkeit:

$$B(n,p,k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}. \quad (4)$$

Der Erwartungswert der Binomialverteilung berechnet sich durch:

$$\lambda = n \cdot p \quad (5)$$

Die Varianz der Binomialverteilung berechnet sich wie folgt:

$$\sigma^2 = Var(X) = n \cdot p \cdot (1-p). \quad (6)$$

Eine Binomialverteilung mit den Parametern n und p lässt sich durch eine Normalverteilung annähern, falls gilt (*Laplace-Bedingung*):

$$\sqrt{n \cdot p \cdot (1-p)} > 3. \quad (7)$$

Bei einer Bernoulli-Kette ist es oftmals spannend zu wissen, wie hoch die Wahrscheinlichkeit ist, dass der erste Treffer nach k Versuchen auftritt. Die Verteilungsfunktion hierzu bezeichnen wir als *geometrische Verteilung*:

$$P(k) = (1-p)^{k-1}p \tag{8}$$

Wenn die Wahrscheinlichkeit P eines Treffers bei wachsemden n gegen 0 geht[1] und gleichzeitig $\lim_{n\to\infty} = n \cdot p = \lambda > 0$ gilt, lassen sich die binomialverteilten Wahrscheinlichkeiten bei wachsemden n mit der *Poisson-Verteilung* approximieren:

[1] $\lim_{n\to\infty} n = 0$.

$$P(X=k) = \frac{\lambda^k e^{-\lambda}}{k!} \quad k = 0, 1, .. \quad \lambda > 0 \tag{9}$$

Stetige Verteilungen

Die Dichtefunktion[2] einer stetigen Verteilung zeichnet sich, wie der Name impliziert durch das Vorhandensein eines Wert an allen Stellen aus [3]. Wir nutzen hier anstelle einer Verteilungsfunktion eine Dichtefunktion [4]:

[2] Die Dichtefunktion ist das *stetige* Gegenstück zur *diskreten* Verteilungsfunktion

[3] Der Definitionsbereich ist $\mathbb{R}$.

[4] Das Integral bis x stellt die Wahrscheinlichkeit $P(X \leq x)$ dar.

$$\int_{-\infty}^{x} f(t)dt \tag{10}$$

Damit es sich um eine Dichtefunktion einer Verteilung handelt, müssen zwei Eigenschaften erfüllt sein:

$$f(t) \geq 0 \quad \forall t \in \mathbb{R} \tag{11}$$

$$\int_{-\infty}^{\infty} f(t)dt = 1 \tag{12}$$

(11): Wahrscheinlichkeiten können nicht negativ sein
(12): Insgesamt beträgt die kumulierte Wahrscheinlichkeit 1

Gleichverteilung

Bei einer Gleichverteilung definieren wir die Dichtefunktion:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{für} \quad x \in [a,b] \\ 0 & \text{sonst} \end{cases} \tag{13}$$

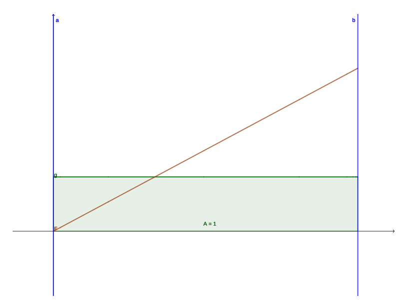

Abbildung 1: Beispiel einer Gleichverteilung

Die Wahrscheinlichkeiten berechnen sich:

$$P(X \leq x) = F(X) = \begin{cases} 0 & \text{für} \quad x < a \\ \frac{x-a}{b-a} & \text{für} \quad a \leq x \leq b \\ 1 & \text{für} \quad x > b \end{cases} \tag{14}$$

Exponentielle Verteilung

Die Dichtefunktion einer exponentiellen Verteilung hat die Form:

$$f(x) = \lambda \cdot e^{-\lambda x} \tag{15}$$

Nach Bedingung (12) müssen wir zeigen, dass $\int_{-\infty}^{\infty} = 1$.[5]

[5] Da wir mit einer Exponentialfunktion arbeiten, die wir für $x \geq 0$ definieren, beweisen wir analog $\int_0^{\infty} = 1$.

Beweis.

$$\int_0^{\infty} f(t)dt = \lambda \int_0^{\infty} e^{-\lambda t} dt = \lambda \cdot [\frac{e^{-\lambda t}}{-\lambda}]_0^{\infty}$$
$$= [-e^{-\lambda t}]_0^{\infty} = -0 - (-1) = 1$$

□

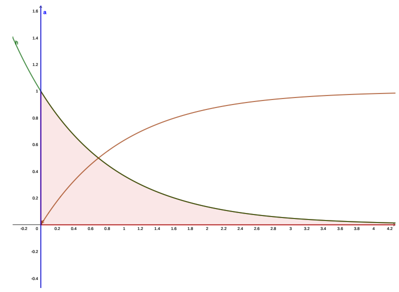

Abbildung 2: Beispiel einer Exponentialverteilung

Die entsprechende Wahrscheinlichkeitsfunktion:

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{für} \quad x > 0 \\ 0 & \text{für} \quad x \leq 0 \end{cases} \tag{16}$$

Normalverteilung

Die Normalverteilung folgt der grundlegenden Form $f(t) = e^{-t^2}$. Überprüfen wir zunächst die Eigenschaften (11)[6] und (12).

[6] trivial, $e^0 = 1$

Zu (12):

$$\int_{-\infty}^{\infty} e^{\frac{-t^2}{2}} dt = \sqrt{2\Pi}$$

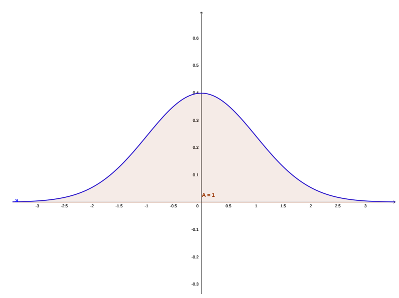

Abbildung 3: Beispiel einer Normalverteilung

Da $\sqrt{2\Pi} > 1$, ist Bedingung (12) nicht erfüllt. Wir definieren daher die *Standard-Normalverteilung* ϕ wie folgt [7]:

[7] Wir normieren die Funktion auf das berechnete Integral, damit $\int_{-\infty}^{\infty} = 1$

$$\phi(t) = \frac{e^{\frac{-t^2}{2}}}{\sqrt{2\Pi}} \tag{17}$$

Zusätzlich definieren wir die *allgemeine Normalverteilung* [8]:

[8] μ: Verschiebung entlang der x-Achse
σ: Stauchung/Streckung

$$\phi(x) = \frac{1}{\sqrt{2\Pi}\sigma} \cdot e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \tag{18}$$

Die Wahrscheinlichkeitsverteilung berechnet sich durch[9]:

[9] Die Wahrscheinlichkeitsverteilung zu einer Dichtefunktion ist einfach das Integral

$$\Phi(x) = \frac{1}{\sqrt{2\Pi}\sigma} \int_{-\infty}^{x} e^{-\frac{1}{2}(\frac{t-\mu}{\sigma})^2} dt \tag{19}$$

Hierbei wird ein uneigentliches Integral[10] verwendet, dessen Berechnung nur über Näherungsverfahren möglich ist. Hierfür gibt es Tabellen, die in der Klausur gestellt werden.

[10] Integral mit ∞ als Grenze, Wert nicht einfach berechenbar

Unabhängigkeit zweier Zufallsvariablen

Sei (Ω, P) ein Wahrscheinlichkeitsraum und $X, Y : \Omega \mapsto \mathbb{R}$ zwei Zufallsgrößen. X und Y heißen *unabhängig*, wenn $\forall x \in W(X), y \in W(Y)$:

$$P(X = x, Y = y) = P(X = x) \cdot P(Y = y) \tag{20}$$

gilt. Bei gemeinsamen Verteilungen sprechen wir von *diskreten* Zufallsvariablen. Die Unabhängigkeit lässt sich zeigen durch eine Vierfeldertafel, bei der alle Werte durch Multiplikation berechnet werden und eine Addition der Einzelwahrscheinlichkeiten 1 ergibt (siehe Beispiel 2). Die Abhängigkeit lässt sich durch ein Gegenbeispiel beweisen.

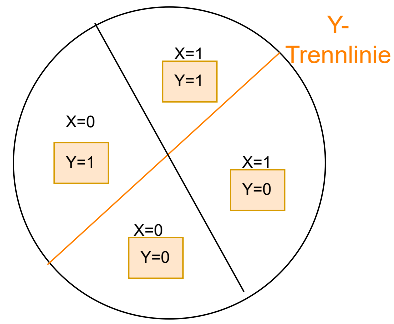

Abbildung 4: Mögliche Kombinationen zweier Zufallsvariablen in Ω.

Beispiel 2: Beispiel einer unabhängigen, gemeinsamen Verteilung. Hier sehen wir das an der Vierfeldertafel[11]:

X/Y	Y=0	Y=1	
X=0	$\frac{1}{16}$	$\frac{3}{16}$	0.25
X=1	$\frac{3}{16}$	$\frac{9}{16}$	0.75
	0.25	0.75	1

[11] Siehe Beispielsweise: $P(X = 0, Y = 0) = P(X) \cdot P(Y) = \frac{1}{4} \cdot \frac{1}{4} = \frac{1}{16}$

Die Wahrscheinlichkeitsverteilung einer diskreten 2-dimensionalen Zufallsgröße lässt sich durch die Wahrscheinlichkeitsfunktion

$$f(x) = \begin{cases} P_{ik} & \text{falls} \quad x = x_i, y = y_k \\ 0 & \text{sonst} \end{cases} \tag{21}$$

darstellen.

Erwartungswert

Für eine *diskrete* Zufallsgröße ist der Erwartungswert

$$E(X) = \sum_i x_i P(X = x_i). \tag{22}$$

Beispiel 3: Erwartungswert einer gegebenen Wahrscheinlichkeitsverteilung

x_i	0	2	3	4	10	11
p_i	$\frac{3}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$

$$E(X) = 0 \cdot \frac{3}{8} + 2 \cdot \frac{1}{8} + 3 \cdot \frac{1}{8} + 4 \cdot \frac{1}{8} + 10 \cdot \frac{1}{8} + 11 \cdot \frac{1}{8}$$

Für eine *stetige* Zufallsgröße ist der Erwartungswert [12]:

$$E(X) = \int_{-\infty}^{\infty} y \cdot f(y) dy \tag{23}$$

[12] Funktioniert prinzipiell genau wie im *diskreten*, aber mit Integral statt Summe

Beispiel 4: Erwartungswert einer stetigen Verteilung[13]

Gegeben eine Dichtefunktion f:

$$f(x) = \begin{cases} ax^2, 0 \leq x \leq 5 \\ 0, \text{sonst} \end{cases}$$

Wir berechnen zuerst a, sodass die Bedingungen (11) und (12) erfüllt sind:

$$\int_{-\infty}^{\infty} f(x)dx = \int_0^5 ax^2 dx = 1$$
$$[\frac{ax^3}{3}]_0^5 = \frac{125a}{3} - \frac{0a}{3} = \frac{125a}{3} = 1$$
$$a = \frac{3}{125}$$

Der Erwartungswert berechnet sich nun wie folgt:

$$E(X) = \int_0^5 x \frac{3}{125} x^2 dx$$
$$= [\frac{3}{125} \cdot \frac{x^4}{4}]_0^5 = \frac{3}{125} \cdot 6254 = \frac{15}{4}$$

[13] gerne in Klausuren gefragt

Rechenregeln für den Erwartungswert

Sei (Ω, P) ein Wahrscheinlichkeitsraum und $X, Y : \Omega \mapsto \mathbb{R}$ zwei Zufallsgrößen. Es gilt:

$$E(X+Y) = E(X) + E(Y) \tag{24}$$
$$E(a \cdot X + b \cdot Y) = a \cdot E(X) + b \cdot E(Y) \tag{25}$$
$$\text{Sind X , Y unabhängig dann gilt} \quad E(X \cdot Y) = E(X) \cdot E(Y) \tag{26}$$

Varianz

Der mittlere Abstand der Werte vom Erwartungswert. Die quadratische Funktion ist besser als der Betrag analytisch, weshalb wir diese verwenden.

Die Varianz zum Beispiel 3:

$$Var(X) = -\frac{15}{4}^2 \frac{3}{8} + \frac{-7}{4}^2 \frac{1}{8} + \frac{-3}{4}^2 \frac{1}{8} + \frac{1}{4}^2 \frac{1}{8} + \frac{25}{4}^2 \frac{1}{8} + \frac{28}{4}^2 \frac{1}{8}$$

Die Varianz zum Beispiel 4:

$$f(x) = \begin{cases} \frac{3}{125}x^2, 0 \le x \le 5 \\ 0, \text{sonst} \end{cases}$$
$$E(X) = \frac{15}{4}$$
$$Var(X) = \int_{-\infty}^{\infty} (x - \frac{15}{4})^2 f(x)dx$$
$$= \int_0^5 (x - \frac{15}{4})^2 \frac{3}{125}x^2 dx$$

Die Varianz lässt sich so (einfacher) auch so berechnen:[14]

$$Var(X) = E(X^2) - (E(X))^2 \tag{27}$$

[14] Das ist nicht besser, aber einfacher. Nicht die schöne Definition, sondern den schnellen Weg wählen.

Beweis.

$$\begin{aligned} Var(X) :&= E((X - E(X))^2) \\ &= E(X^2 - 2E(X) \cdot X + (E(X))^2) \\ &= E(X^2) - E(2E(X) \cdot X) + E((E(X))^2) \\ &= E(X^2) - E(X)E(X) + (E(X))^2 \\ &= E(X^2) - (E(X))^2 \end{aligned}$$

□

x_i	0	2	3	4	10	11
p_i	$\frac{3}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$
x_i^2	0	4	9	16	100	121

aktualisierte Tabelle zu Beispiel 3

Die Varianz zum Beispiel 3:

$$Var(X) = \frac{250}{8} - \frac{15^2}{4^2} = \frac{500 - 225}{16} = \frac{225}{16}$$

Die Varianz zum Beispiel 4:

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx = \frac{3}{125}\int_0^5 x^2 x^2 dx$$
$$= \frac{3}{125}[\frac{x^5}{5}]_0^5 = 15.$$

$$Var(X) = 15 - (\frac{15}{4})^2 = \frac{15 \cdot 16 - 15^2}{16} = \frac{15}{16}.$$

Die Ungleichung von Tschebyscheff

$M = W(A) \cap (] - \infty; E(A) - a) \cup (E(A) + a; +\infty[)$

$Var(A) := \sum_{i=1}^{k}(x_i - E(A))^2 \cdot P(A = x_i) \ge \sum_{x_i \in M}(x_i - E(A))^2 P(A = x_i) \ge \sum_{x_i \in M} a^2 P(A = x_i) = a^2 \sum_{x_i \in M} P(A = x_i) = P(|A - E(A)| \ge a)$

Anwenden bei: symmetrischem Intervall um Erwartungswert herum, Erwartungswert und Varianz müssen vorhanden sein. Frage: im symmetrischen Intervall oder außerhalb des symmetrischen Intervalls.

Beispiel 5: Obsthändler, Binomialverteilung, $n = 240$, $p = \frac{1}{20}$. Gesucht $P(6 \leq X \leq 18)$? $E(X) = np = 240 \cdot \frac{1}{20} = 12$ $Var(X) = 12 \cdot 1920 = \frac{57}{5} \approx 11.4$ Das Intervall ist geeignet für Tschebyscheff, weil $E(X)$ genau in der Mitte des Intervalls liegt. $P(6 \leq X \leq 18) = 1 - P(|X - 12| \geq 7) \geq 1 - \frac{11.4}{7^2} \approx 1 - 0.2326 = 0.7673$. Es liegen *mindestens* 76% im Intervall.

Satz

Sei $X_1, ..., X_n$ eine Folge identisch verteilter unabhäniger Zufallsgrößen mit $E(X_i) = \mu$ und $Var(X_i) = \sigma^2 \forall i \in \mathbb{N}$,

$$U_n = \frac{(X_1 + ... + X_n) - n\mu}{\sqrt{n} \cdot \sigma}. \tag{28}$$

Dann konvergieren die Verteilungsfunktionen F_n der Zufallsgrößen U_n gegen die Standardnormalverteilung.

Beispiel 6:

$$\begin{aligned} P(6 \leq X \leq 18) &= P(\frac{6-12}{\sqrt{11.4}} \leq \frac{X-12}{\sqrt{11.4}} \leq \frac{18-12}{\sqrt{11.4}} \\ &= \Phi(\frac{6}{\sqrt{11.4}}) - \Phi(\frac{-6}{\sqrt{11.4}}) \\ &= 2\Phi(\frac{6}{\sqrt{11.4}}) - 1 \approx 0.9232 \end{aligned}$$

Hinweis: $P(6 \leq X \leq 18) = P(5 < X < 19)$.

Wenn die Varianz nicht zu groß ist, kann ein Korrekturfaktor (0.5) bei der Berechnung der Approximation durch die Standardnormalverteilung aufaddiert werden (Siehe Folie 94).

Bernoullisches Gesetz der großen Zahlen

Satz: Sei $n \in \mathbb{N}, 0 < p < 1, v_n = \frac{A}{n}$ die relative Häufigkeit eines Ereignisses A in einer Bernoullikette der Länge n. Es gelte außerdem $P(A) = p$. Dann gilt für jedes $\epsilon > 0$

$$\lim_{n \to +\infty} P(|v_n - p| > \epsilon) = 0. \tag{29}$$

ϵ ist die Ungenauigkeit.

Beispiel 7: Bernoullikette der Länge n mit der Trefferwahrscheinlichkeit p, A binomialverteilt, neue Zufallsgröße: $\frac{A}{n}$

$E(\frac{A}{n}) = \frac{1}{n} E(A) = \frac{1}{n} \cdot n \cdot p = p$ $Var(\frac{A}{n}) = \frac{1}{n^2}$i

$P(|\frac{A}{n} - p| > \epsilon) \leq \frac{p\cdot(1-p)}{n\cdot\epsilon^2}$ Geht für n gegen ∞ gegen 0.

Beispiel 8: Eine Münze wird 10000 mal geworfen, Wahrscheinlichkeit für 4990 bis 5010 mal Kopf.$n = 10000, p = \frac{1}{2}, E(X) = 5000, Var(X) = 2500$

$$P(4990 \leq X \leq 5010) = 1 - P(|X - 5000| \geq 11) \tag{30}$$
$$= 1 - \frac{2500}{11^2} = 1 - \frac{2500}{121} \approx -19.6 \tag{31}$$

Nach der Ungleichung von Tschebyscheff wissen wir, dass die Wahrscheinlichkeit größer als -19.6 ist, korrekt aber naja.

Mit der Standardnormalverteilung:

$$\begin{aligned} P(&\frac{4990-5000}{50} \leq \frac{X-5000}{50} \leq \frac{5010}{50}) \\ &= P(-\frac{1}{5} \leq T \leq \frac{1}{5}) \\ &\approx \Phi(\frac{1}{5}) - \Phi(-\frac{1}{5}) \\ &= 2\Phi(\frac{1}{5}) - 5 \approx 0.1586 \end{aligned}$$

Die Kovarianz von zwei Zufallsgrößen

Die *Kovarianz* von zwei Zufallsgrößen X und Y wird durch

$$cov(X, Y) = E((X - E(X)))(Y - E(Y)) \tag{32}$$

Wenn die Kovarianz nicht 0 ist, sind die Variablen *immer* abhängig.

Statistik

Methodensammlung zum Umgang mit Wahrscheinlichkeiten/Zahlen in der realen Welt.

Grundgesamtheit: Gesamtheit gleichartiger Objekte, die hinsichtlich eines bestimmten Merkmals untersucht werden sollen.

Eine *Stichprobe* vom Umfang $n \in \mathbb{N}$ ist eine Folge $X_1, ..., X_n$ von unabhängigen identisch verteilten Zufallsgrößen. Eine *Statistik* ist eine Zufallsgröße

$$g(X_1, ..., X_n) \text{ mit } g : \mathbb{R}^n \mapsto \mathbb{R}. \tag{33}$$

Der *Modus* ist der häufigste Wert, der in der Stichprobe vorkommt.

Der *Median* ist der Wert, der genau in der Mitte der Datenverteilung liegt. Bei einer geraden Anzahl von Individualdaten ist die Hälfte der Summe der beiden in der Mitte liegenden Werte.

Beispiel 9: Altersverteilung Studierende
Stichprobe: 23, 25, 22, 23, 36, 23, 20, 18, 19, 20, 24
Modus: 23, da der Wert dreimal vorkommt.
Geordnet (hervorg: Median): 18, 19, 20, 20, 22, 23, 23, 23, 24, 25, 26
Mittelwert: 23

Erwartungstreuer Schätzer für den Erwartungswert

Sei $(X_1, ..., X_n)$ eine Stichprobe vom Umfang n mit $E(X_i) = \mu \in \mathbb{R}$ und $Var(X_i) = \sigma^2$ für alle $i \in \{1, ..., n\}$ und

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

Es gilt:

$$E(\bar{X}) = \mu \tag{34}$$

$$Var(\bar{X}) = \frac{\sigma^2}{n} \tag{35}$$

Man sagt, $\bar{X}$ ist ein erwartungstreuer Schätzer für μ.
Herleitung für (35): $Var(\bar{X}) = (\frac{1}{n})^2 \sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2} \cdot n \cdot \sigma^2 = \frac{\sigma^2}{n}$.

Varianz einer Stichprobe

Sei $(X_1, ..X_n)$ eine Stichprobe vom Umfang n mit $E(X_i) = \mu \in \mathbb{R}$ und $Var(X_i) = \sigma^2 in \mathbb{R}^+$ für alle $i \in 1, ..., n$ und

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \tag{36}$$

mit $\bar{X} =$
Herteilung einer alternativen Formel für die Varianz

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i^2 - 2\bar{X}X_i + (\bar{X})^2) \tag{37}$$

$$= \frac{1}{n-1}\sum_{i=1}^{n} X_i^2 - 2\bar{X}\sum_{i=1}^{n} X_i + n(\bar{X})^2 \tag{38}$$

$$= \frac{1}{n-1}\sum_{i=1}^{n} X_i^2 - 2(\bar{X})^2 \cdot n + n(\bar{X})^2 \tag{39}$$

$$= \frac{1}{n-1}(\sum_{i=1}^{n} X_i^2 - n \cdot (\bar{X})^2) \tag{40}$$

Beispiel 10: Kennzahlen für Lineare Regression berechnen
1) $\bar{X}, \bar{Y}$
2)
3)