Regressionsrechnung (Übungen)

1. **Bei der Schulärztin**

   a) 10 Buben aus einer 3. Klasse Volksschule wurden gemessen und gewogen:

Größe in cm	137	131,5	136,5	131,5	141,5	130,5	130	139	138	134
Gewicht in kg	31,5	25,5	32	24	37	26,5	27,5	31,5	35	27

   Untersuche, ob zwischen Größe (x) und Gewicht (y) ein linearer Zusammenhang besteht. Ermittle die Gleichung der Regressionsgeraden und den Korrelationskoeffizienten und interpretiere das Ergebnis.

   b) Dasselbe für 8 Mädchen aus derselben Klasse:

Größe in cm	132,5	135	145	131	127,5	126,5	125,5	135,5
Gewicht in kg	32,5	31	37,5	29,5	25	23,5	29	37

2. **Durchschnittstemperatur**

   Von folgenden österreichen Städten sind die Seehöhe und die Jahresdurchschnittstemperatur in einem bestimmten Zeitraum bekannt:

Stadt	Seehöhe (x)	Temperatur (y)
Wien	203 m	9,1° C
Salzburg	437 m	8,6° C
Innsbruck	579 m	8,4° C
Graz	342 m	9,4° C
Klagenfurt	448 m	8,1° C

   a) Ermittle die Gleichung der Regressionsgeraden und den Korrelationskoeffizienten. Interpretiere das Vorzeichen der Steigung.

   b) Welche Durchschnittstemperatur ist für einen Ort in 1000 m Seehöhe zu erwarten?

   c) In welcher Höhe müsste ein Ort liegen, an dem die Durchschnittstemperatur 10° C beträgt?

3. **Blutdruck**

   Herr Gross misst eine Woche lang jeden Tag seinen Blutdruck. Er erhält folgende Werte: 135/90, 128/85, 144/98, 150/102, 145/95, 156/112, 136/90. (Die erste Zahl gibt den systolischen (maximalen) Druck, die zweite den diastolischen (minimalen) Druck in mmHg an.) Er will untersuchen, ob es zwischen dem ersten und zweiten Wert einen linearen Zusammenhang gibt.

a) Ermittle die Gleichung der Regressionsgeraden (x: systolischer Druck, y: diastolischer Druck) und den Korrelationskoeffizienten.

b) Interpretiere den Korrelationskoeffizienten in diesem Sachzusammenhang.

4. **Zuckerbelastung**

Mit einer Zuckerbelastung kann man feststellen, ob jemand in Gefahr ist, Diabetes zu bekommen. Dabei wird nach einer süßen Mahlzeit in regelmäßigen Abstanden der Blutzuckergehalt gemessen. Bei einer Patientin erhielt man folgende Werte:

Zeit in Stunden	0,5	1,0	1,5	2,0	2,5	3,0
Blutzucker in mg/dl	210	195	160	135	130	115

a) Ermittle die Gleichung der linearen Regressionsfunktion Z(t) (t: Zeit in h, Z(t): Zuckergehalt in mg/dl) und interpretiere ihre Steigung.

b) Welcher Blutzuckerwert ist nach 4 h zu erwarten?

5. **Geburten**

In Österreich gab es in den vergangenen Jahren die folgenden Geburtenzahlen (in Tausend):

Jahr	1950	1960	1970	1980	1990	2000	2010
Geburten (in 1000)	107,5	125,9	112,3	90,9	90,4	78,3	78,7

a) Zeichne die Werte in ein Koordinatensystem und ermittle die Gleichung der Regressionsgeraden (t: Zeit in Jahren seit 1950, N(t): Geburtenzahl in 1000)

b) Erstelle aufgrund dieses Modells eine Schätzung für das Jahr 2020.

c) In Wirklichkeit wurden 2020 ca. 83600 Kinder geboren. Erkläre anhand der Grafik, warum die Schätzung ungenau ist.

6. **Luftemperatur in Wien**

Die Tabelle zeigt die Jahresdurchschnitte der Lufttemperatur in Wien.
(Quelle: Statistisches Jahrbuch der Stadt Wien 2020)

Jahr	1960	1970	1980	1990	2000	2010	2020
Temperatur	9,5 °C	9,2 °C	8,7 °C	10,9 °C	11,7 °C	9,9 °C	11,9 °C

a) Berechne die Gleichung der linearen Regressionsfunktion T(t) (t: Zeit in Jahren seit 1960, T(t): Temperatur in °C) und den Korrelationskoeffizienten.

b) Interpretiere den Wert der Steigung der Regressionsgeraden im Sachzusammenhang.

c) Erstelle mit diesem Modell eine Prognose für das Jahr 2050.

7. **Zehnkampf**

Bei den olympischen Spielen 2021 in Tokio fanden Wttkämpfe in Zehnkampf statt. Die folgende Tabelle zeigt die Leistungen aller Sportler, die es in die Endwertung geschafft haben, in den Disziplinen 100 m-Lauf, Weitsprung und 1500 m-Lauf.

Name	Land	100 m-Lauf Zeit in s	Weitsprung Weite in m	1500 m-Lauf Zeit in min
Steven Bastien	USA	10,69	7,39	4:26,95
Felipe Dos Santos	BRA	10,58	7,38	4:52,40
Cedric Dubler	AUS	10,89	7,36	5:03,69
Johannes Erm	EST	11,04	7,36	4:28,42
Adam S. Helcelet	CZE	11,06	7,16	4:44,74
Kai Kazmirek	GER	11,09	7,48	4:48,30
Pierce Lepage	CAN	10,43	7,65	4:31,85
Kevin Mayer	FRA	10,68	7,50	4:43,17
Ashley Moloney	AUS	10,34	7,64	4:39,19
Martin Roe	NOR	10,86	7,03	4:47,58
Garrett Scantling	USA	10,67	7,30	4:35,54
Ilja Schkurenjow	ROC	10,93	7,59	4:34,62
Vitali Schuk	BLR	11,04	6,93	4:42,57
Jiri Sykora	CZE	11,18	7,03	4:54,97
Karel Tilga	EST	11,31	6,77	4:38,28
Marcel Uibo	EST	11,32	7,37	4:38,64
Jorge Ureña	ESP	10,66	7,30	4:27,82
Lindon Victor	GRN	10,67	7,24	4:54,32
Damian Warner	CAN	10,12	8,24	4:31,08
Pawel Wiesiolek	POL	10,83	7,27	4:30,02
Zach Ziemek	USA	10,55	7,20	4:38,38

a) Ist ein guter Sprinter auch ein guter Springer? Berechne den Korrelationskoeffienten zwischen den Zeiten im 100 m-Lauf und den Weiten beim Weitsprung und interpretiere ihn.

b) Ist ein guter Sprinter auch ein guter Mittelstreckenläufer? Berechne den Korrelationskoeffienten zwischen den Zeiten im 100 m-Lauf und im 1500 m-Lauf und interpretiere ihn.

8. **Hunde**

Von 10 Hunden verschiedener Rassen wurde die Schulterhöhe in cm (x) und das Gewicht in kg (y) gemessen.

Rasse	x	y	Rasse	x	y
Berner Sennenhund	66	40	Franz. Bulldogge	32	12
Chihuahua	22	3	Irish Setter	65	28
Chow-Chow	48	24	Mops	28	8
Dackel	23	9	Pudel	40	15
Deutscher Schäfer	60	35	Husky	55	22

a) Ermittle die Gleichung der Regressionsgeraden und den Korrelationskoeffizienten. Argumentiere, ob die lineare Regression in diesem Fall ein gutes Modell darstellt.

b) Schätze mit diesem Modell das Gewicht eines Collies mit 50 cm Schulterhöhe ab.

9. **Strafen im Straßenverkehr**

In der Tabelle sind die Strafen in Euro für Alkohol am Steuer (x) und Geschwindigkeitsüberschreitung um 20 km/h (y) für einige EU-Länder zusammengestellt. (Quelle: www.bussgeldrechner.org)

Land	Alkohol	Schnellfahren	Land	Alkohol	Schnellfahren
BE	170	100	AT	300	30
BG	255	25	PL	145	25
DE	500	35	PT	250	60
FR	135	135	RO	280	60
GR	200	100	SK	200	35
IE	200	80	SI	300	80
IT	530	35	ES	500	100
HR	405	135	CZ	100	40
LU	145	50	HU	370	110
NL	325	185	CY	100	35

a) Stelle die Daten in einem Koordinatensystem dar (Streudiagramm).

b) Ermittle die Gleichung der linearen Regressionsfunktion und den Korrelationskoeffizienten.

c) Erkläre anhand des Streudiagramms, warum der Korrelationskoeffizient so niedrig ist und was das bedeutet.

10. **Schulbesuch und Kinderzahl**

Die nachstehende Tabelle zeigt für ausgewählte Länder die durchschnittliche Dauer des Schulbesuchs (in Jahren) von Frauen ab 25 Jahren und die mittlere Kinderzahl pro Frau im Jahr 2009. (Quelle: www.gapminder.org/tools)

Land	Schulbesuch (x)	Kinder (y)
Deutschland	12,0	1,38
Frankreich	10,5	1,99
Russland	12,9	1,54
USA	13,7	2,00
Brasilien	7,2	1,82
Argentinien	10,1	2,38
Kolumbien	6,4	2,04
Japan	12,2	1,36
Iran	5,1	1,77
Bangladesh	2,6	2,32
Äthiopien	1,0	5,06
DR Kongo	4.0	6,52
Nigeria	4,1	5,87
Australien	11,5	1,94

Ermittle die Gleichung der Regressionsgeraden und interpretiere ihre Steigung im Sachzusammenhang.

11. **Gebrauchtwagen**

Ein Auto einer bestimmten Marke kostet neu 29900 €. Die Preisentwicklung ist in der folgenden Tabelle ersichtlich.

Alter in Jahren	1	2	3	5	10	15	20
Preis in €	24900	21900	18900	12900	6900	2900	900

a) Argumentiere, warum ein exponentielles Regressionsmodell hier besser passt als ein lineares.

b) Erstelle eine exponentielle Regressionsfunktion in der Form $p(t) = a \cdot b^t$ (t: Zeit in Jahre, p(t): Preis).

c) Berechne, in welcher Zeit sich nach diesem Modell der Preis halbiert.

12. **Pkw-Bestand**

Die Tabelle zeigt die Anzahl der in Österreich zugelassenen Pkw und wie viele davon mit Diesel betrieben werden/wurden (in Tausend, gerundet).

(Quelle: www.statistik.at)

Jahr	Pkw	Diesel
1960	400	10
1970	1200	37
1980	2250	80
1990	3000	410
1995	3600	830
2000	4100	1500
2005	4200	2130
2010	4400	2450
2015	4750	2700
2020	5100	2760

a) Erstelle eine lineare Regressionsfunktion $N_1(t)$, die die Anzahl aller Pkw beschreibt, und berechne den Korrelationskoeffizienten.
(t: Zeit in Jahren seit 1960, N_1: Gesamtzahl der Pkw in Tausend.)

b)

- Argumentiere, warum die Anzahl der Diesel-Pkw von 1960 bis 2005 besser durch eine Exponentialfunktion beschrieben werden kann.
- Ermittle die exponentielle Regressionsfunktion $N_2(t) = a \cdot b^x$ und interpretiere den Wert des Parameters b.
(t: Zeit in Jahren seit 1960, N_2: Anzahl der Diesel-Pkw in Tausend.)
- Welche Prognose ergibt sich nach diesem Modell für 2020? Warum ist diese Prognose nicht realistisch?

c) In einem anderen Modell soll die Anzahl der Diesel-Pkw von 2000 bis 2020 durch eine quadratische Funktion angenäherit werden. Ermittle die Gleichung der quadratischen Regressionsfunktion N_3.
(t: Zeit in Jahren seit 2000, N_3: Anzahl der Diesel-Pkw in Tausend.)

Ergebnisse

1.
   a) $y = 0,97 \cdot x - 101,60$; $r = 0,908$
   b) $y = 0,67 \cdot x - 58,03$; $r = 0,846$

   In beiden Fällen starker positiver Zusammenhang (je größer, desto schwerer).

2.
   a) $y = -0,00264 \cdot x + 9,78$; $r = -0,7$; Je höher der Ort liegt, umso kälter ist es.
   b) 7,1°C
   c) –82,6 m (!)

   (Aufgrund der wenigen Daten kann man keine genaue Aussage machen!)

3.
   a) $y = 0,91 \cdot x - 33,44$; $r = 0,969$
   b) Da r nahe bei 1 liegt, besteht ein starker positiver linearer Zusammenhang zwischen dystolischem und diastolischem Blutdruck.

4.
   a) $Z(t) = -39,7 \cdot t + 227$
      Der Blutzuckergehalt nimmt pro Stunde um durchschnittlich 39,7 mg/dl ab.
   b) 68 mg/dl

5.
   a) $N(t) = -0,731 \cdot t + 119,7$
   b) 68,5
   c) Die Geburtenzahlen sind zuerst gestiegen, danach stark gesunken und am Ende wieder leicht gestiegen. Daher kann man sie nicht gut durch eine lineare Funktion annähern.

6.
   a) $T(t) = 0,04 \cdot t + 9$; $r = 0,714$
   b) Die Temperatur pro Jahr um durchschnittlich 0,04 °C zu.
   c) $T(90) = 12,7$ °C

7.
   a) $r = -0,688$; mittlerer negativer Zusammenhang (je kürzer die Laufzeit, umso länger die Sprungweite)
   b) $r = 0,246$; schwacher positiver Zusammenhang (je kürzer die Laufzeit auf 100 m, umso kürzer ist auch die Laufzeit auf 1500 m)

8.

a) $y = 0{,}68 \cdot x - 10{,}1$; $r = 0{,}95$

Die lineare Regression ist ein gutes Modell, weil r nahe bei 1 liegt.

b) 23,7 kg

9.

a)

b) $y = 0{,}046 \cdot x + 60{,}19$; $r = 0{,}139$

c) Die Punkte sind weit verstreut und lassen sich nicht gut durch eine Gerade annähern. Zwischen der Höhe der Strafen für Alkohol am Steuer und Geschwindigkeitsüberschreitung besteht fast kein Zusammenhang.

10. $y = -0{,}27 \cdot x + 4{,}51$

Pro zusätzlichem Jahr Schulbesuch bekommt eine Frau durchschnittlich um 0,27 Kinder weniger.

11.

a) Der Preis nimmt zuerst schnell, dann immer langsamer ab.

Der Preis kann nicht negativ werden.

b) $p(t) = 31052 \cdot 0{,}845^t$

c) 4,13 Jahre

12.

a) $N_1(t) = 79{,}12 \cdot x + 570; \quad r = 0{,}992$

b) $N_2(t) = 9{,}64 \cdot 1{,}13^t$; die Anzahl der Diesel-Pkw nahm pro Jahr um 13 % zu

$N_2(60) \approx 15500$ (das wären 15,5 Millionen!); die Anzahl der Diesel-Pkw kann nicht größer werden als die Gesamtzahl aller Pkw.

c) $N_3(t) = -3{,}475 \cdot t^2 + 130{,}9 \cdot t + 1517$