2. Statistik mit zwei abhängigen Variablen

2.1. Der Korrelationskoeffizient

1. **Beispiele**

   Es gibt relativ viele Beispiele von Abhängigkeiten zwischen zwei Messgrössen.

   a) Grössere Menschen sind tendenziell schwerer.

   b) Die Jahres-Niederschlagsmenge einer Messstation ist abhängig von der Höhe (über Meeresspiegel).

   c) Die durchschnittliche Lebenserwartung von Menschen ist abhängig vom Bildungsniveau.

   d) usw.

   In diesem Kapitel geht es darum, den Zusammenhang zwischen den jeweiligen Variablen zu beschreiben und zu berechnen. Dabei beschränken wir uns auf die Fälle, bei denen die Messgrössen in Zahlwerten vorliegen. Das dritte Beispiel in der obigen Aufzählung lassen wir also unbetrachtet.

2. **Musterbeispiel**

   Wir betrachten zwei abhängige Messreihen:

x_i	2	4	5	7
y_i	1	4	2	5

3. **Definition**

. . . .

. . . .

. . . .

. . . .

. . . .

4. **Was beschreibt der Korrelationskoeffizient?**
Um diese Frage zu beantworten, betrachten wir vier Beispiele:

a) $(x_i \,|\, y_i) = (1|1), (2|3), (4|4), (6|6),$

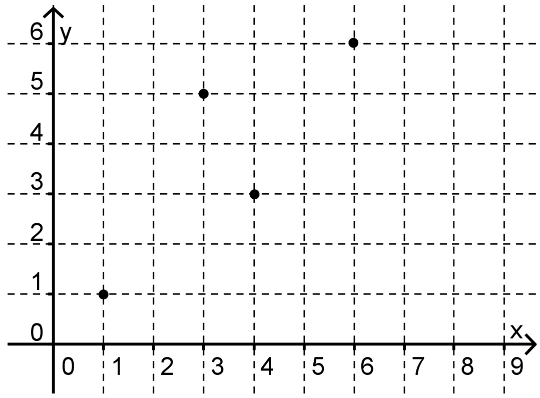
b) $(x_i \,|\, y_i) = (1|2), (3|1), (4|6), (6|4),$

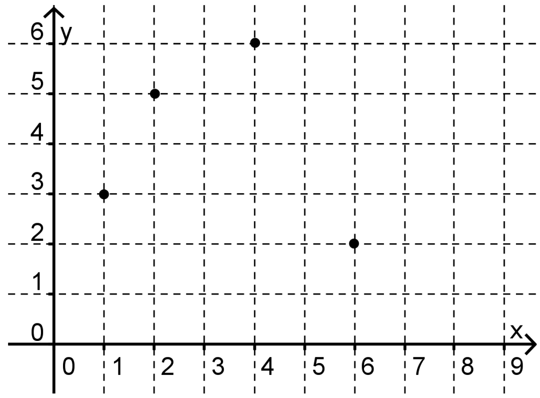
c) $(x_i \,|\, y_i) = (1|6), (2|4), (4|2), (6|1),$

d) $(x_i \,|\, y_i) = (1|5), (2|1), (5|6), (6|2),$

5. **Grafisches**

Bestimme den Korrelationskoeffizienten anhand der Grafiken.





Übung

Man nehme sich zwei (verschiedenfarbige) Würfel und werfe beide zusammen viermal. Der eine Würfel bezeichnet die x-Koordinate eines Punktes, der andere die y-Koordinate. So erhält man 4 Punkte. Bestimme den Korrelationskoeffizienten dieser Punktverteilung.

2.2. Lineare Regression

1. **Bemerkung**

   Im folgenden geht es darum, durch eine Menge von Punkten im Koordinatensystem die optimale Gerade zu legen. Dabei stellt sich dann die Frage, was unter der optimalen Geraden zu verstehen ist.

2. **Musterbeispiel**

   Wir betrachten die Messpunkte $(1\,|\,1)$, $(4\,|\,2)$ und $(10\,|\,6)$

3. **Übung**

Bestimme die Regressionsgerade durch $(0|8)$, $(2|3)$, $(3|1)$ und $(5|-2)$.

4. **Übung**

Ebenso: $(1|1)$, $(2|-1)$, $(3|0)$, $(4|1)$.

5. **Überlegungsaufgaben**

Was meinst du zu den folgenden Beispielen von Korrelation und Regression?

a) In einer Zeitschrift wurde ein Artikel publiziert, in welchem Daten von vielen Ehepaaren verglichen wurden. Das Alter des Ehemanns korreliert mit dem Alter der Ehefrau.

....

....

....

b) Aus dem gleichen Artikel: Die Körpergrösse des Ehemanns korreliert mit der Grösse der Ehefrau.

....

....

....

c) Und noch eine dritte Situation: Das Körpergewicht des Ehemanns korreliert mit dem Körpergewicht der Ehefrau.

....

....

....

6. **Bringt der Storch die Kinder?**

Jemand hat festgestellt, dass die Population der Störche mit der Anzahl Geburten in Zusammenhang gebracht werden kann. Man spricht von einer Scheinkorrelation.

Jahr	1972	1973	1974	1975	1976
Storchenpaare	130	138	150	126	115
Geburten($\times 1000$)	82	86	87	83	78

2.3. Regressionskurven

1. **Bemerkung**

   In der Praxis ist die gesuchte, bestmöglich an die Daten angepasste Kurve, wohl eher selten eine Gerade. Je nach Aufgabenstellung kann es sich um eine Polynomfunktion, Potenz- oder Exponentialfunktion handeln. Beispiele sind: Wurfparabel (Polynomfunktion), Volumenberechnungen, Gasgesetz etc. (Potenzen) und Radioaktivität (Exponentialfunktion).

2. **Wurfparabel**

   In einem Experiment wird eine Kugel durch die Luft geworfen. Dabei darf angenommen werden, dass die Flugbahn der Kugel eine Parabel $y = f(t) = a \cdot t^2 + b \cdot t + c$ beschreibt.
   Mit Hilfe von Lichtschranken in verschiedenen Höhen misst man die Durchgangszeiten der Kugel.

Schranke	1.5 m	2 m	2.5 m
Zeiten	$0.09, 1.03$	$0.21, 0.88$	$0.42, 0.68$

   Die Lichtschranke in 1.5 m Höhe stoppte die Kugel nach 0.09 Sekunden (im Flug nach oben) und nach 1.03 Sekunden (im Flug nach unten).

   a) Führe eine quadratische Regression durch. Welches ist die Gleichung des bestangepassten Funktion?

   b) Beantworte folgende Fragen:

   b_1) Zu welcher Zeit t landet die Kugel am Boden?

   b_2) Wie hoch war die Abwurfhöhe?

   b_3) Welche maximale Höhe erreichte die Kugel? Wann war das der Fall?

3. Vergleich von Regressionskurven

Gegeben sind die folgenden Messwerte:

x_i	3	5	10	16	25
y_i	20	10	4	2	1

Welche Regressionskurve passt am besten durch diese Messpunkte?

4. Sättigungskurve

Aus einem chemischen Experiment erhält man die folgenden Daten:

Zeit	0	5	10	15	20	30	40	60
Prozent	5	22	35	45.5	53	64	70	76.5