

## Zugänge zur linearen Regression

---

**Karel Tschacher**

### **Kurzfassung des Inhalts:**

Es wird ein Weg aufgezeigt, wie Schülerinnen und Schülern ohne Kenntnisse der Differenzialrechnung an die Gaußsche Methode der minimalen Fehlerquadrate herangeführt werden können. Die mathematischen Hintergründe der linearen Regression werden dann – für Lehrerinnen und Lehrer – als Hintergrundwissen dargestellt.

### **Klassenstufe(n):**

Die Aufgaben eignen sich sowohl für Klasse 6 ohne Verwendung des Funktionsbegriffs und deren Graphen, als auch für Klasse 8 bei Kenntnissen über Funktionen und deren Graphen.

### **Lernziele:**

Die Schülerinnen und Schüler sollen ...

- bei Alltagssituationen proportionale Zusammenhänge erkennen und diese quantitativ beschreiben;
- Einsicht in verschiedene Verfahren gewinnen, die dazu dienen, Zusammenhänge bei gegebenen Daten zu beschreiben;
- unterschiedliche Darstellungen zur Bearbeitung von metrischen zweidimensionalen Datenlisten verwenden können.

### **Vorkenntnisse bezüglich der Bedienung des Graphikrechners:**

Es sind keine Vorkenntnisse für die Nutzung des Graphikrechners erforderlich.

### **Zeitbedarf:**

Jeder Arbeitsauftrag kann im Rahmen einer Doppelstunde behandelt werden.

### **Sonstige Materialien:**

Zu den Arbeitsaufträgen sind beispielhafte Daten in den Lösungsvorschlägen zu finden. Damit die Schülerinnen und Schüler diese Daten selbst ermitteln können, werden folgende Materialien benötigt:

Lineal, Bandmaß sowie für

- Arbeitsauftrag 1: Runde Alltagsgegenstände
- Arbeitsauftrag 2: Schulhefte
- Arbeitsauftrag 3: Seite eines Fahrtenbuches eines Autofahrers
- Arbeitsauftrag 4: „Ihr Zugbegleiter“ eines ICE

## Begleittext

---

Die drei Grunderfahrungen, die den Mathematikunterricht nach Winter (1996) als allgemeinbildend herausstellen, liegen auch den KMK-Standards zugrunde:

1. *„Erscheinungen der Welt um uns, die uns alle angehen oder angehen sollten, aus Natur, Gesellschaft und Kultur, in einer spezifischen Art wahrzunehmen und zu verstehen.“*

In dem hier vorgeschlagenen Unterrichtsentwurf sollen Schülerinnen und Schüler bei Alltagsgegenständen bzw. -situationen, z. B. bei Gebrauchsgegenständen, Schulheften oder Fahrtstrecken, proportionale Zusammenhänge konstruieren, entdecken oder erkennen.

2. *„Mathematische Gegenstände und Sachverhalte, repräsentiert in Sprache, Symbolen, Bildern, und Formeln, als geistige Schöpfung, als eine deduktiv geordnete Welt eigener Art kennen zu lernen und zu begreifen.“*

Schülerinnen und Schüler sollen die Vorgehensweise einer statistischen Auswertung von Alltagsdaten kennenlernen. Dabei werden sie mit Arbeitsweisen und Verfahren vertraut, die eigene Fachbegriffe und Darstellungsformen erfordern. Sie entwickeln Vorstellungen über die Struktur einer empirischen oder statistischen Untersuchung und lernen Begriffe der Statistik kennen. Sie verwenden dabei den ClassPad als ein Werkzeug zur Problemlösung.

3. *„In der Auseinandersetzung mit Aufgaben Problemlösefähigkeiten, die über die Mathematik hinaus gehen, (heuristische Fähigkeiten) zu erwerben.“*

In diesem Unterrichtsentwurf sollen die Schülerinnen und Schüler Fragestellungen eigenständig aufwerfen, Vorgehensweisen selbst finden und die Güte von Ergebnissen einschätzen lernen.

Die Unterrichtssequenz geht der Frage nach, wie man Beziehungen zwischen Daten von zwei gegebenen Größen erkennen kann. Dabei werden nur proportionale Zusammenhänge betrachtet, damit die Minimierungsaufgabe der Abstände zwischen Messwerten und Schätzwerten eindimensional bleibt. Die allgemeine lineare Regression erfolgt mit einer linearen Funktion, so dass diese Minimierung der Abstände zu einer Aufgabe mit zwei Variablen wird. Das fordert dann allerdings eine mehrdimensionale Statistik, was die Inhalte der Schule übersteigt. Daher werden hier nur Ausgleichsgeraden durch den Ursprung gesucht, der Regressionskoeffizient ist dann die Steigung der Geraden. Der Begriff „Korrelationskoeffizient“ (das Maß für die Güte einer linearen Anpassung der Messwerte) wird nicht eingeführt, die Abweichungen werden nur qualitativ beobachtet und bewertet.

Im Rahmen dieses Artikels werden einfache Begriffe der Statistik angesprochen:

- die graphische Auswertung von Daten mit zwei Merkmalen,
- die Interpretation einer Punktwolke,
- die funktionale Abhängigkeit zweier Merkmale,

- Einführen einer Schätzfunktion zur Vorhersage eines Wertepaares,
- die Güte oder die Fehler einer Schätzfunktion,
- Methoden, Fehler klein zu halten.

Bei dieser Unterrichtssequenz dient der ClassPad nicht nur zur Vereinfachung von Rechnungen, sondern vor allem zur Veranschaulichung der Daten und dem Finden der Regressionsfunktion.

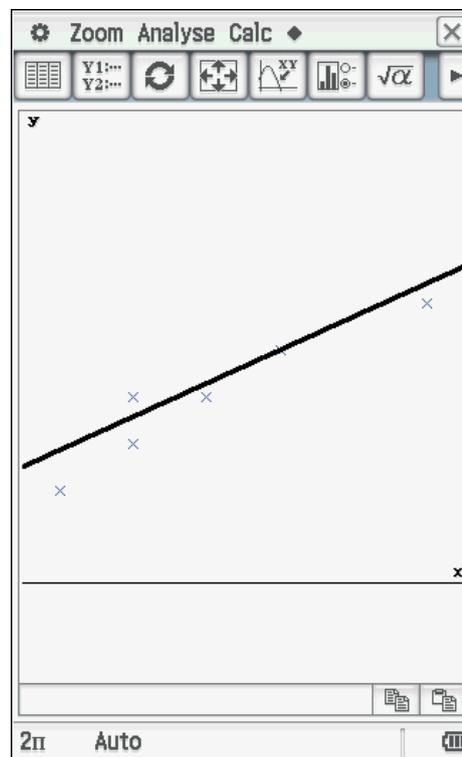
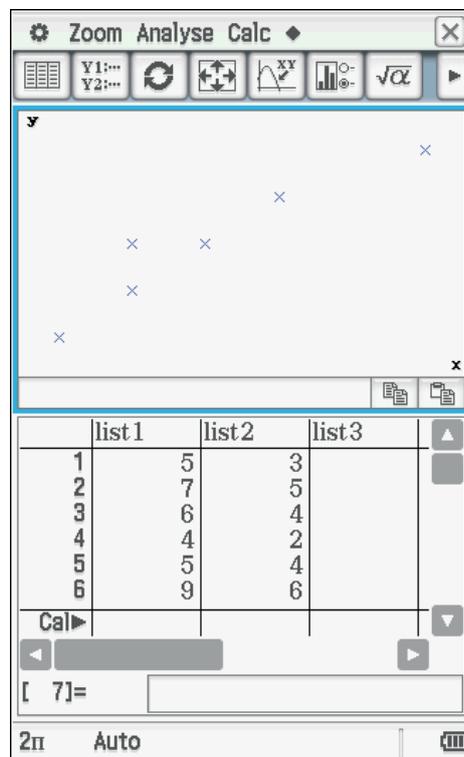
Zunächst wird der mathematische Zusammenhang an einem Beispiel erläutert.

## Mathematische Herleitung

Eine Aufgabe als Einstieg:

„Für ein Modegeschäft wird in einem Zeitraum die Zahl  $x$  der vor dem Schaufenster stehenden Personen und die Zahl  $y$  der anschließend das Geschäft betretenden Kunden beobachtet.“

Zahl $x$ der die Auslagen ansehenden Personen	5	7	6	4	5	9
Zahl $y$ der den Laden betretenden Personen	3	5	4	2	4	6



In den Rechner werden die Datenpaare im Statistikmenü eingegeben, die Punktwolke erzeugt (*Grafik einst* → *Einstellung* → ) und die Gerade der sog. „Schätzfunktion“ gezeichnet (*Calc* → *Regressionen* → *lineare Regression*), einer Funktion, die den Zusammen-

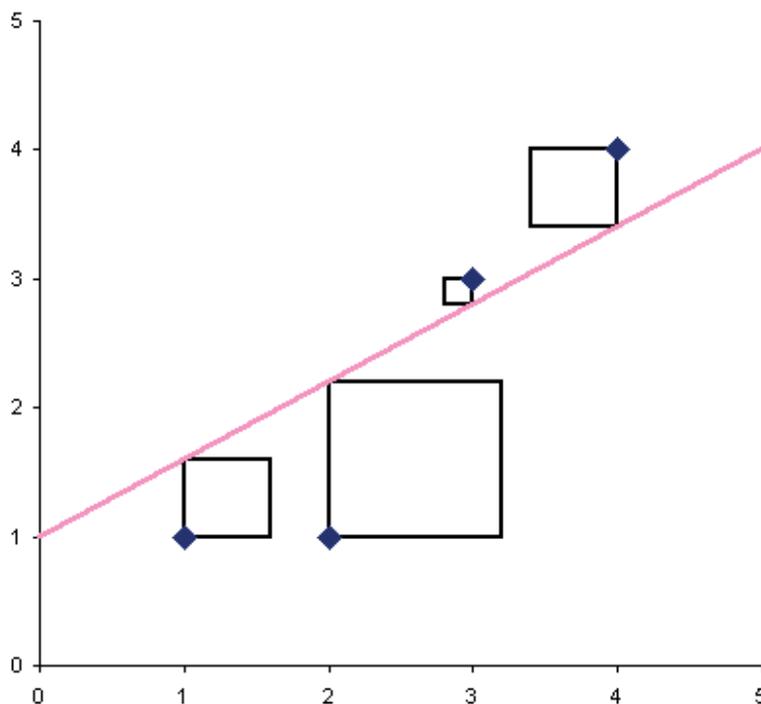
hang zwischen den beiden betrachteten Merkmalen beschreibt oder beschreiben könnte. Die Eingaben in den ClassPad sind am Ende des Artikels dargestellt.

Wie erhält man diese Funktion bzw. Gerade?

Hierzu könnte man sich verschiedene Methoden denken:

- Man könnte die Summe der Abstände der Punkte von der Geraden oder
- die Differenz der  $y$ -Werte der gegebenen Punkte  $P_i(x_i, y_i)$  und den Punkten  $P_g(x_i, y_g)$  auf der Geraden minimieren.

Die auf Carl Friedrich Gauß zurückgehende Idee besteht darin, die Quadrate der Differenz der Punkte entsprechend b) zu verwenden. Geometrisch lässt sich das so veranschaulichen, dass über den  $y$ -Abstandsstrecken Quadrate gelegt werden, wobei ein Eckpunkt der entsprechende Messpunkt  $(x, y)$  ist und der andere auf der Geraden liegt. Die Summe der Quadratflächeninhalte soll minimal sein. Dann wird diese Gerade als diejenige angesehen, die den Zusammenhang der Datenpaare am besten beschreibt.



Zunächst werden im Folgenden Schülerarbeitsaufträge formuliert, die die eigenständige Erarbeitung der linearen Regression ermöglichen. Dann folgen Lösungshinweise zu diesen Arbeitsaufträgen. Die theoretische Herleitung für die Berechnung der Regressionsgeraden findet sich – als Hintergrundwissen für interessierte Lehrerinnen und Lehrer – am Ende des Artikels.

## Schülerarbeitsaufträge

---

### Aufgabe 1

#### **Der Zusammenhang von Durchmesser und Umfang bei Kreisen**

Die Schülerinnen und Schüler verwenden ein Lineal und ein Bandmaß. Sie messen den Durchmesser und den Umfang von Alltagsgegenständen und tabellarisieren die Wertepaare.

Im Schülergespräch werden Fragen aufgeworfen, denen dann im Unterricht nachgegangen werden kann.

### Aufgabe 2

#### **Der Zusammenhang von Länge und Breite bei Schulheften (Din-A-Formate)**

Die Schülerinnen und Schüler messen die Längen und Breiten ihrer Schulhefte, vom Vokabelheft bis zum Zeichenblock und tabellarisieren die Wertepaare.

### Aufgabe 3

#### **Zusammenhang von Benzinpreis, Benzinmenge und zurückgelegter Fahrstrecke**

Die Schülerinnen und Schüler erhalten eine Seite eines Fahrtenbuches eines Autofahrers für den Monat April. Alternativ werden die Eltern gebeten, Daten der Autofahrten zu notieren, dann werden die Daten in der Klasse zusammengetragen und ausgewertet.

### Aufgabe 4

#### **Zusammenhang Fahrzeit und Strecke bei einem ICE der Bundesbahn**

Die Schülerinnen und Schüler erhalten das Blatt „Ihr Zugbegleiter“ eines ICE, aus dem sie selbst gewählte Daten entnehmen.

## Lösungshilfen zu Aufgabe 1: Kreiszahl Pi

---

Die Schülerinnen und Schüler messen den Durchmesser und den Umfang verschiedener runder Gegenstände und erstellen eine Liste. Im Folgenden wurden beispielhafte Messwerte verwendet.

	Durchmesser ( <i>cm</i> )	Umfang ( <i>cm</i> )
Tasse	7,9	25,3
Milchkanne	10,3	32,1
Untertasse	13,8	44,6
Kuchenteller	18	57,1
Brotteller	20	61,5

Diese Aufgabe kann sowohl in der 6. als auch in der 8. Klassenstufe bearbeitet werden. Die Schülerinnen und Schüler erkennen unmittelbar, dass mit dem Durchmesser auch die Umfänge zunehmen. Mögliche Fragen, die bei der Bearbeitung der Aufgaben auftreten könnten, sind etwa die Folgenden:

### Klassenstufe 6

- Wie kann man den Zusammenhang beschreiben?
- Wie genau sind diese Messungen?
- Kann man den Umfang aus dem Durchmesser berechnen?
- Wie kann man bei bekanntem Durchmesser den Umfang eines Gegenstandes vorher-sagen?
- Mit welchem Fehler ist bei der Vorhersage zu rechnen?

### Klassenstufe 8

- Wie kann man die Kreiszahl  $\pi$  aus den Daten ermitteln?
- Wie kann der Rechner bei diesen Berechnungen helfen?
- Kann man einen Term (eine Funktionsgleichung) für die Vorhersage finden, um bei bekanntem Durchmesser den Umfang zu berechnen?
- Welche Eigenschaften der Funktion kann man aus dem Graphen erkennen?

Die gemessenen Daten werden im Menü Statistik  eingegeben. Dazu wird *list1* mit „Durchm“ und *list2* mit „Umfang“ überschrieben. Die Eingabe der Daten erfolgt zellenweise und jede Eingabe wird mit der **EXE**-Taste abgeschlossen (s. Abb. 1).

Im Statistik-Menü kann die „Punktewolke“ dargestellt werden. Dazu wird zunächst das Graphikfenster festgelegt. Mit dem Befehl *Grafik einst*  → *Einstellung* wird der *StatGraph1* eingestellt. Die Listennamen findet man am Ende der Darstellung von *X-List* und *Y-List*. Hier kann man nach Wunsch andere Darstellungen wählen. Man beendet die Eingabe mit *Einst* oder Verlassen der Seite (s. Abb. 2). Dann kehrt man auf die Seite der Listen zurück und wählt das Menü *Graphen anzeigen*  (s. Abb. 3). Mit dem Befehl *Verfolgen* im Menü *Analyse* (alternativ: ) kann man die Wertepaare durchlaufen.

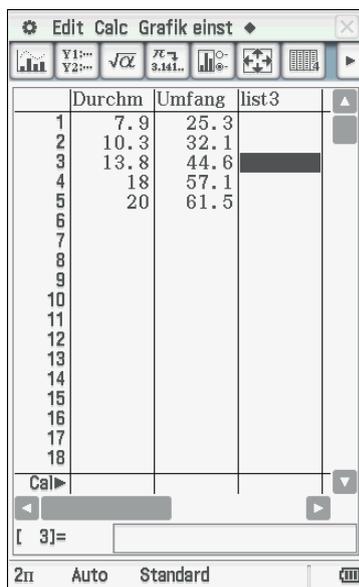


Abb. 1: Eingabe von Daten in Listen

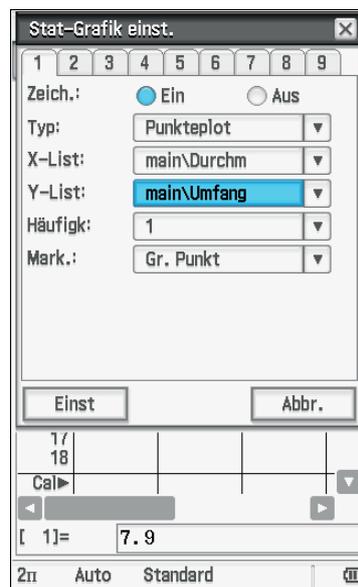


Abb. 2: Einstellung

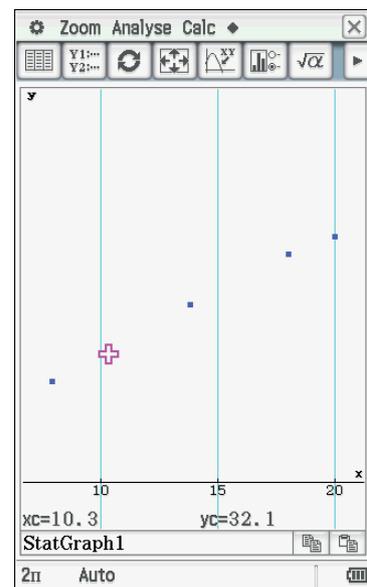


Abb. 3: Die Punktewolke

Die Quotienten der Wertepaare können bestimmt werden, daraus lässt sich dann die Kreiszahl  $\pi$  berechnen. Dazu wechselt man in das Main-Menü . Man schreibt den gewünschten Bruch mit den Namen der Listen und ordnet der Liste der Brüche einen Namen zu, den man zuvor in die erste Zeile von *list3* geschrieben hat, hier der Name „Kreis“ für Kreiszahl. Die Werte sind oft als Brüche mit größeren Zählern und Nennern dargestellt. Ein Wechsel zu Dezimalzahlen erfolgt mit dem Befehl Näherungswert . Zum Zuordnen einer Liste oder eines Ergebnisses verwendet man die Speichertaste. Dazu verwendet man die Tastatur **Keyboard** und dann in der ersten Zeile am Ende den Pfeil  $\Rightarrow$  (s. Abb. 4). Der Mittelwert dieser fünf Quotienten wird gebildet. Der Befehl für Mittelwert heißt *mean*. Man findet ihn im Menü *Aktion* → *Liste* → *Statistik* (oder Eingabe der Buchstaben mithilfe der Tastatur). Den Namen der Liste kann man auch durch Kopieren (im Menü *Statistik* → *Einfügen* → *Kopieren*) im Main-Menü erzeugen. Hier wird der Wert der Variablen „p“ zugeordnet (s. Abb. 5). Der Mittelwert von der Kreiszahl ist  $p = 3,159628556$ , mit der eingestellten Genauigkeit von neun Nachkommastellen. Die Schätzwerte des Umfangs können mit der Rechnung:  $Umfang = p \cdot Durchmesser$ , kurz

$Uschaetz = p \cdot Durchm$  ermittelt werden. Dazu wird *list4* in "Uschaetz" umbenannt und im Main-Menü wird die Rechnung und der Speicherplatz eingegeben (s. Abb. 6).

	Durchm	Umfang	Kreisz
1	7.9	25.3	3.2025
2	10.3	32.1	3.1165
3	13.8	44.6	3.2319
4	18	57.1	3.1722
5	20	61.5	3.075

list= main \ Kreisz

Umfang / Durchm → Kreisz  
 {3.202531646, 3.116504854}

Abb. 4: Die Quotienten der Wertepaare

	Durchm	Umfang	Kreisz
1	7.9	25.3	3.2025
2	10.3	32.1	3.1165
3	13.8	44.6	3.2319
4	18	57.1	3.1722
5	20	61.5	3.075

list= main \ Kreisz

mean (Kreisz) → p  
 3.159628556

Abb. 5: Der Mittelwert von der Kreiszahl

	Umfang	Kreisz	Uschaetz
1	25.3	3.2025	24.961
2	32.1	3.1165	32.544
3	44.6	3.2319	43.603
4	57.1	3.1722	56.873
5	61.5	3.075	63.193

list= main \ Kreisz

Abb. 6: Die Schätzwerte mit der Kreiszahl p

Die Abweichungen der Schätzwerte von den Messwerten werden berechnet, um die Güte der Schätzung zu ermitteln. Dazu wird die nächste Liste mit „Abweich“ beschriftet, die Rechnung in das Main-Menü eingegeben und in der Liste „Abweich“ abgespeichert (s. Abb. 7). Die Summe dieser Werte sollte möglichst „klein“, d. h. möglichst Null oder nahe Null sein. Dazu wird die Summe der Einzelwerte der Liste „Abweich“ gebildet. Im Main-Menü schreibt man *sum* und den Namen der Liste in runden Klammern dazu, dann werden die Listenwerte addiert. Man kann den Befehl *sum* auch unter *Aktion* → *List* → *Berechnungen* finden. Die Abweichung ist  $-0,57$  (s. Abb. 8). Ist der Näherungswert für die Kreiszahl „p“ besser, wenn man die Mittelwerte der Umfänge durch die Mittelwerte der Durchmesser dividiert? Der Wert  $p_2 = 3,151428571$  ist wenig verschieden von  $p = 3,159628556$ . Ist damit auch der Fehler bei den Schätzwerten kleiner? Dazu berechnet man den Bruch aus den Mittelwerten und speichert das Ergebnis unter „p2“. Dabei zeigt sich, dass es einen Unterschied zwischen der Kreiszahl „p“ aus der ersten Berechnung und der Kreiszahl „p2“ aus der letzten Berechnung gibt, wenn der Unterschied auch klein ist (s. Abb. 9).

Abbildung 7 zeigt den Rechner in der 'Edit'-Ansicht für die Liste 'main \ Abweich'. Die Tabelle enthält folgende Werte:

Kreis	Uschaetz	Abweich
1	3.2025	0.3389
2	3.1165	-0.444
3	3.2319	0.9971
4	3.1722	0.2267
5	3.075	-1.693
6		

Die untere Anzeige zeigt die Liste: `Umfang-Uschaetz→Abweich` mit den Werten `{0.3389344074, -0.4441741}`.

Abb. 7: Die Liste der Abweichungen der Einzelwerte des Umfangs vom Schätzwert

Abbildung 8 zeigt den Rechner mit dem Befehl `sum (Abweich)`. Die Anzeige zeigt den Wert `-0.5739989218`.

Abb. 8: Summe der Einzelwerte der Liste „Abweich“

Abbildung 9 zeigt den Rechner mit dem Befehl `mean (Umfang) / mean (Durchm)`. Die Anzeige zeigt den Wert `3.151428571`. Darunter ist `p-p2` mit dem Wert `0.00819998459` zu sehen.

Abb. 9: Differenz von p und p2

Wenn wir nun die Summe der Abweichungen von den neu berechneten Schätzwerten „Abweich2“ betrachten, sehen wir, dass nun die Abweichungssumme tatsächlich Null ist (s. Abb. 10). Das ist erst einmal ein Ergebnis, das man sich gewünscht hat. Die Schätzfunktion passt sich den Messwerten gut an.

Es lässt sich fragen, ob es andere Möglichkeiten für eine „gute“ Annäherung gibt. Betrachten wir die Beträge der Abweichungen, diese sollen möglichst klein werden. Zunächst führen wir diese Überlegung für „p“, dem Mittelwert der Quotienten, durch. Man rechnet die Beträge der Abweichungen aus und bildet die Summe dieser nichtnegativen Zahlen. Den Befehl für den Betrag der Werte findet man in der Tastatur **Keyboard** **Math1** → **|■|** oder man schreibt mit Buchstaben das Wort *abs* und dann den Namen der Liste in runden Klammern. Zuvor hat man die nächste Liste mit dem Namen „Betr1“ beschriftet. Die Summe der Beträge der Abweichungen wird berechnet. Die Summe aller Fehler ist nun 3,6994 (s. Abb. 11). Wie sieht diese Summe bei „p2“ aus? Dazu errechnet man die Schätzwerte mit „p2“ und bestimmt die Summe der Beträge der Abweichungen und nennt das „Betr2“. Der Fehler ist da noch größer (s. Abb. 12).

Abweich | Schaetz... | Abweic...

1	0.3389	24.896	-0.404
2	-0.444	32.46	0.3597
3	0.9971	43.49	-1.11
4	0.2267	56.726	-0.374
5	-1.693	63.029	1.5286
6			

list= main\Abweich2

Schaetz2-Umfang→Abweich2  
{-0.4037142857, 0.3597142}

sum (Abweich2)  
0

Abb. 10: Summe der Abweichungen von den neu berechneten Schätzwerten

Schaetz... | Abweic... | Betr1

1	24.896	-0.404	0.3389
2	32.46	0.3597	0.4442
3	43.49	-1.11	0.9971
4	56.726	-0.374	0.2267
5	63.029	1.5286	1.6926
6			

list= main\Betr1

|Abweich|→Betr1  
{0.3389344074, 0.44417412}

sum (Betr1)  
3.699491573

Abb. 11: Summe der Beträge der Abweichungen für „p“

Abweic... | Betr1 | Betr2

1	-0.404	0.3389	
2	0.3597	0.4442	
3	-1.11	0.9971	
4	-0.374	0.2267	
5	1.5286	1.6926	
6			

[ 1 ]=

|Abweich2|→Betr2  
{0.4037142857, 0.35971428}

sum (Betr2)  
3.776571429

Abb. 12: Summe der Beträge der Abweichungen für „p2“

Vielleicht gibt es noch andere Wege der Annäherung? Eine Möglichkeit ist es, da wir ja an nichtnegativen Werten interessiert sind, mit den Quadraten der Abweichungen zu rechnen. Quadrate sind nichtnegativ. Werte kleiner eins werden als Quadrat noch kleiner und Werte größer eins werden größer als die Grundzahl. Das bedeutet, dass man jetzt auch noch die Fehler gewichtet. Allerdings kann man dann nur die Summe der Abweichungsquadrate vergleichen. Den Befehl für das Quadrieren findet man auf der Gerätetaste  $\square^2$  gefolgt von einer 2 für „hoch 2“. Der Name für diese Liste ist „Quadr1“, wenn wir mit „p“ die quadratischen Abweichungen betrachten. Für „p“ ergibt sich eine Summe der quadratischen Abweichungen von 4,222 (s. Abb 13). Nun wiederholt man die Rechnung noch einmal für „p2“. Diese Liste heißt dann „Quadr2“. Hier ist die Summe kleiner als zuvor (Abb. 14).

Betr1 | Betr2 | Quadr1

1	0.3389	0.4037	0.1149
2	0.4442	0.3597	0.1973
3	0.9971	1.1103	0.9943
4	0.2267	0.3743	0.0514
5	1.6926	1.5286	2.8648
6			

list= main\Quadr1

(Abweich)^2→Quadr1  
{0.1148765325, 0.19729065}

sum (Quadr1)  
4.222610838

Abb. 13: Summe der Abweichungsquadrate für „p“

Abweic... | Betr1 | Betr2

1	-0.404	0.3389	0.4037
2	0.3597	0.4442	0.3597
3	-1.11	0.9971	1.1103
4	-0.374	0.2267	0.3743
5	1.5286	1.6926	1.5286
6			

list= main\Abweich2

(Abweich2)^2→Quadr2  
{0.1629852245, 0.12939436}

sum (Quadr2)  
4.001734367

Abb. 14: Summe der Abweichungsquadrate für „p2“

	Durchm	Umfang	Kreiszahl	Uschaetz	Abweich	Schaetz2	Abweich2	Betr1	Betr2	Quadr1	Quadr2
1	7.9	25.3	3.2025316	24.961066	0.3389344	24.896286	-0.403714	0.3389344	0.4037143	0.1148765	0.1629852
2	10.3	32.1	3.1165049	32.544174	-0.444174	32.459714	0.3597143	0.4441741	0.3597143	0.1972907	0.1293944
3	13.8	44.6	3.2318841	43.602874	0.9971259	43.489714	-1.110286	0.9971259	1.1102857	0.9942601	1.2327344
4	18	57.1	3.1722222	56.873314	0.226686	56.725714	-0.374286	0.226686	0.3742857	0.0513865	0.1400898
5	20	61.5	3.075	63.192571	-1.692571	63.028571	1.5285714	1.6925711	1.5285714	2.864797	2.3365306
6											
7											
8											
9											
10											
11											
12											
13											
14											
15											
16											
17											
18											
19											

Abb. 15: Alle Listen auf einen Blick

**Fazit:**

Die Werte von „p“ und „p2“ sind je nach Vergleichsmethode für die Schätzung von Umfängen besser oder schlechter. Kann man einen Wert für die Kreiszahl finden, bei der diese Summe besonders klein, also minimal wird? Wie erhält man diese? Man muss eine Reihe von Berechnungen anstellen, die verschiedene Werte der Kreiszahl zum Bestimmen des Schätzumfanges mit den Messwerten vergleicht. Sinnvolle Werte für die Kreiszahl sind Werte in der Nähe von 3,15.

Für Klasse 6 ist hier eine Grenze erreicht, weil der Termbegriff, die Verwendung von Funktionsterm und Funktionsgraph noch nicht bekannt sind. Man kann nun die anderen Beispiele von dem Schülerarbeitsblatt mit den dargestellten Methoden in Gruppen durcharbeiten lassen, wobei auch noch andere Messdaten verwendet werden können. Mögliche weitere Aufgaben wären: Zusammenhang von Fläche und Einwohnerzahlen von deutschen Städten, Zusammenhang von Preis und Mengen bei Süßigkeiten oder Getränken und vieles mehr, das Schülerinnen und Schüler interessiert.

**Lösungshilfen zu Aufgabe 2: DIN-Formate von Heften**

Die Schülerinnen und Schüler messen die Länge und die Breite von ihren Schulheften und erstellen eine Liste.

Heft	Länge in mm	Breite in mm
Zeichenblock	418	299
A 4 Heft	295	212
A5 Heft	208	149
A6 Heft	144	106
Vokabelheft	107	72

Die Schülerinnen und Schüler haben im ersten Durchgang wichtige Begriffe erarbeitet und genutzt. Beim zweiten Datensatz werden die Fragen schon spezifischer sein können.

Die Überlegungen aus dem 1. Beispiel werden hier nochmals etwas ausführlicher beschrieben, falls jemand sofort mit diesem Beispiel einsteigen möchte.

Nach Eingabe und graphischer Darstellung der Daten (s. Abb. 16) könnte man einen proportionalen Zusammenhang vermuten, also mit Breite  $B$  und Länge  $L$ :  $B = k \cdot L$  wobei  $k < 1$  ist. Ansatz: Man bestimmt eine Größe, die man Schätzfunktion nennt, die man aus Längen und Breiten berechnen kann und bildet die Abweichung des Schätzwerts vom Messwert:  $a(x) = x \cdot \text{Laenge}[1]\text{Breite}[1] + \dots$ . Es wird also eine Funktion  $a(x)$  definiert. Der Befehl *define* wird einfach mit der Tastatur geschrieben. Die Listen „Laenge“ und „Breite“ sind Spalten (Vektoren). Man kann auf die einzelnen Werte dieser Listen zugreifen, wenn man die Nummer des Platzes in eckigen Klammern hinter den Namen der Liste schreibt.  $\text{Laenge}[3]$  bedeutet, dass man die Zahl des dritten Werts der Liste verwenden will, hier die Zahl 208. So ist der ganze Ausdruck aufgebaut. Wenn man andere Werte in die Listen schreibt, berechnet man die neue Aufgabe, ohne die einzelnen Werte immer neu einzutippen. Der Rechner vereinfacht automatisch den langen Term. Die lange Summendarstellung wird vom Rechner ausmultipliziert, die Ausdrücke werden zusammengefasst und vereinfacht:  $a(x) = 1172x - 838$  (s. Abb. 17).

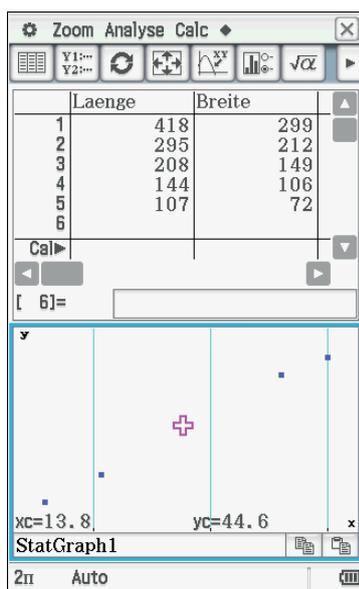


Abb. 16: Die Daten und die Punktwolke



Abb. 17: Abweichung des Schätzwerts vom Messwert

Die Funktion  $a(x)$  wird graphisch dargestellt. Für positive Werte von  $a(x)$ , also Punkte die oberhalb der  $x$ -Achse liegen, bedeutet es, dass die Abweichungen größer als Null sind und für Punkte, die unterhalb der  $x$ -Achse liegen, dass die Abweichungen negativ sind. Im Schnittpunkt mit der  $x$ -Achse sind die Abweichungen Null, diese Nullstelle bestimmt den Wert, bei der die Abweichung am geringsten ist. Die Abweichung ist bei  $x = 0,715$  minimal, nämlich Null (s. Abb. 18). Die Liste „Abweich“ zeigt aber, dass es negative und positive Abweichungen gibt. Das erklärt, weshalb die Summe nahezu Null ist, obwohl alle Werte Abweichungen haben, die von Null verschieden sind. Die Abweichungen werden in eine Liste eingetragen (s. Abb. 19).

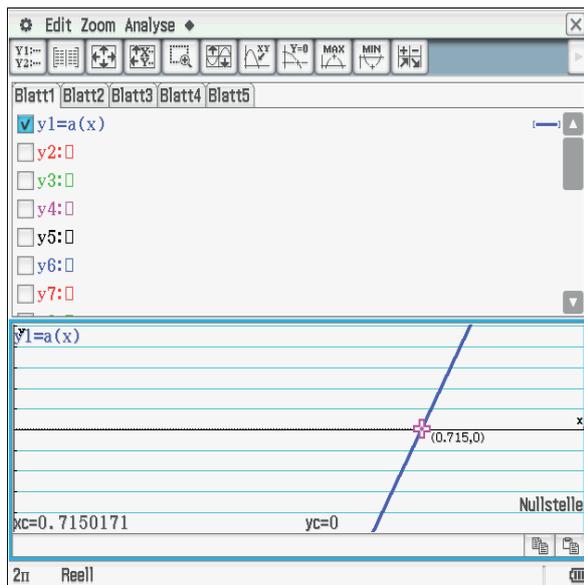


Abb. 18: Schnittpunkt mit der x-Achse

	Laenge	Breite	Abweich
1	418	299	-0.122852
2	295	212	-1.069956
3	208	149	-0.276443
4	144	106	-3.037538
5	107	72	4.5068297
6			
7			

Laenge\*0.7150171-Breite

{-0.1228522, -1.0699555, -0.2764432, -3.037538}

Abb. 19: Liste der Abweichungen

Eine andere Methode, um Abweichungen besser zu erfassen, besteht darin, die Beträge der Abweichungen zu betrachten. Dies ergibt eine abschnittsweise lineare Funktion, das Minimum des Graphen liefert den Proportionalitätsfaktor mit der geringsten betragsmäßigen Abweichung (s. Abb. 22). Anstelle der langen Darstellung der Summe kann eine Summenfunktion verwendet werden. Alle Summanden sind ähnlich aufgebaut. Es wird jedes Mal aus den beiden Listen die gleiche Stelle [i] verwendet. Der Befehl wird über **Keyboard** → **Math2** →  $\sum$  eingegeben.

Das Summenzeichen ist folgendermaßen erklärt:

$$\sum_{i=1}^5 i = 1 + 2 + 3 + 4 + 5;$$

$$\sum_{i=1}^5 3 \cdot i = 3 \cdot 1 + 3 \cdot 2 + 3 \cdot 3 + 3 \cdot 4 + 3 \cdot 5.$$

Die folgenden einfachen Aufgabenbeispiele könnten zum Üben der Summenschreibweise und der Eingabe in den ClassPad dienen: Versuche einmal die ersten zehn natürlichen Zahlen zu addieren:  $1 + 2 + \dots + 10$  (s. Abb. 20). Dann könnte die Gauß-Aufgabe betrachtet werden, bei der die Summe der ersten 100 natürlichen Zahlen berechnet werden soll:  $1 + 2 + 3 + \dots + 99 + 100$ . Zudem könnte auch die Summe der ersten  $n$  natürlichen Zahlen  $1 + 2 + \dots + n$  berechnet werden (s. Abb. 21).

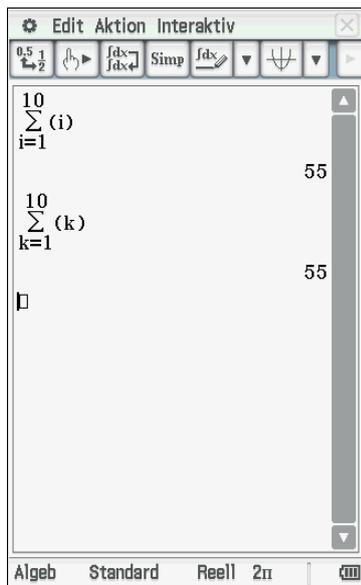


Abb. 20: Eingabe der Summe der ersten zehn natürlichen Zahlen mit Summenzeichen

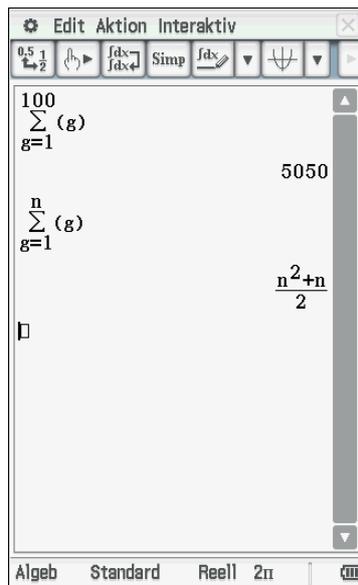


Abb. 21: Eingabe der Summe  $1+2+\dots+99+100$

Zurück zu unserer Frage, die Eingabe sieht dann so aus:

$$\sum_{i=1}^5 (x \cdot \text{Laenge}[i] - \text{Breite}[i]) = x \cdot \text{Laenge}[1] - \text{Breite}[1] + x \cdot \text{Laenge}[2] - \text{Breite}[2] + \dots$$

Das erspart viel Schreibarbeit. Man fügt jetzt noch den Betrag hinzu. Hier wurde der zweite Faktor  $x$  "nach hinten" verlegt. Das Minimum der Betragsabweichungsfunktion liegt bei  $0,7163462$  (s. Abb. 23).

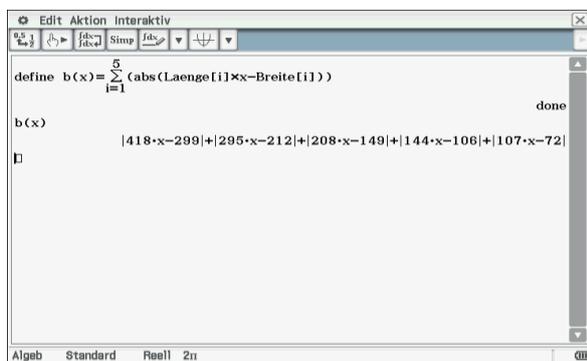


Abb. 22: Die Betragsabweichungsfunktion  $b(x)$

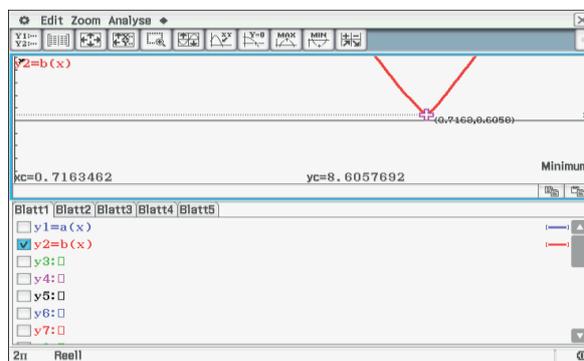


Abb. 23: Der Graph der Betragsabweichungsfunktion  $b(x)$

Die Summe der Abweichungen ist  $1,5577$  (s. Abb. 24).

Eine weitere Methode, um nichtnegative Abweichungen zu erhalten ist, dass man die Abweichungen quadriert und das Minimum dieser Parabeln sucht (s. Abb. 25). Der Proportionalitätsfaktor ist nun  $0,716143$  (s. Abb. 26). Die Summe der quadratischen Abweichungen ist  $1,31$  (s. Abb. 27).

	Laenge	Breite	Abweich	Abweichb
1	418	299	-0.122852	0.4327116
2	295	212	-1.069956	-0.677871
3	208	149	-0.276443	9.6E-6
4	144	106	-3.037538	-2.846147
5	107	72	4.5068297	4.6490434

Laenge×0.7163462-Breite→Abweichb  
 {0.4327116, -0.677871, 0.0000096, -2.8461472, 4.6490434}  
 sum(Abweichb) 1.5577464

Abb. 24: Summe der Abweichungen

$$q(x) = (418-x-299)^2 + (295-x-212)^2 + (208-x-149)^2 + (144-x-106)^2 + (107-x-72)^2$$

Abb. 25: Die Abstandssumme mit quadratischen Abständen  $q(x)$

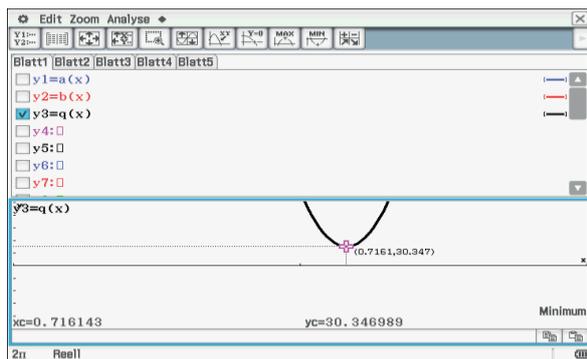


Abb. 26: Graphische Darstellung Abstandsfunktion  $q(x)$

	Laenge	Breite	Abweich	Abweichb	Abweichq
1	418	299	-0.122852	0.4327116	
2	295	212	-1.069956	-0.677871	
3	208	149	-0.276443	9.6E-6	
4	144	106	-3.037538	-2.846147	
5	107	72	4.5068297	4.6490434	

Laenge×0.716143-Breite→Abweichq  
 {0.347774, -0.737815, -0.042256, -2.875408, 4.627301}  
 sum(Abweichq) 1.319596

Abb. 27: Die Summe der quadratischen Abweichungen

Betrachten wir die Geraden der Schätzfunktionen in der Punktelcke. Das ist die Gerade zu  $y_4$ , wenn man die Beträge der Abstände wählt (s. Abb. 28). In Abbildung 29 werden die quadratischen Abstände bei  $y_5$  betrachtet.

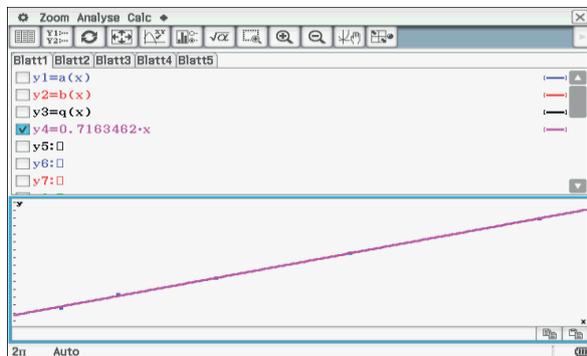


Abb. 28: Graphische Darstellung der Schätzfunktion der Beträge der Abstände

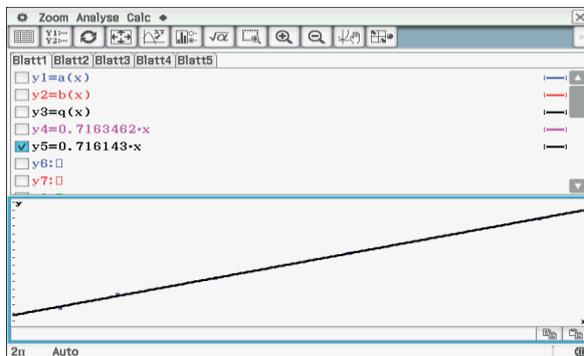


Abb. 29: Graphische Darstellung der Schätzfunktion der quadratischen Abstände

**Fazit:**

Es ergeben sich für diese Datenpaare keine erkennbaren Unterschiede in den Schätzfunktionen. Am Rande angemerkt: Der Proportionalitätsfaktor ist der Kehrwert von Wurzel 2,  $x = \frac{1}{\sqrt{2}} \cong 0,70711$ .

### Lösungshilfen zu Aufgabe 3: Benzinverbrauch eines PKW

Die Schülerinnen und Schüler erhalten eine Seite eines Fahrtenbuches eines Autofahrers für den Monat April. Alternativ werden die Eltern gebeten, Daten der Autofahrten zu notieren, die dann in der Klasse zusammengetragen werden.

Datum	Benzinmenge	Preis	gefahrte Strecke
	Angabe in <i>Liter</i>		Angabe in <i>km</i>
05.04.13	45,23	67,80 €	418
08.04.13	52,89	81,93 €	452
15.04.13	35,89	57,39 €	289
15.04.13	49,26	75,35 €	493
16.04.13	22,45	32,93 €	152
22.04.13	59,57	92,27 €	528
30.04.13	42,49	64,97 €	386

Die Vorgehensweise entspricht weitgehend dem zuvor beschriebenen Vorgehen. Zunächst wird der Zusammenhang Menge und Preis untersucht. Es geht um den Preis pro Liter, der für den Monat April im Mittel gezahlt wurde. Es werden Listen erstellt und der Literpreis je Kauf ermittelt, davon wird dann der Mittelwert *mean* berechnet. Er wird verglichen mit dem Mittelwert aus der Summe der Mengen und der Summe der Rechnungsbeträge (s. Abb. 30). Auch hier ergeben sich unterschiedliche Werte, auch wenn sich das erst an der dritten Stelle hinter dem Komma zeigt (s. Abb. 31).

Mit dem Begriff des arithmetischen Mittels kann man zunächst eine Reihe von Fragen klären. Die Quotienten aus Menge und Preis liefern für jede Tankfüllung die Frage des Tagespreises. Dann kann man von diesen Tagespreisen das arithmetische Mittel bilden. Hier  $k = 1,53162$  €. Liste „LPreis“ ist der Kaufpreis je Tankfüllung, das arithmetische Mittel ist hier  $m = 1,53561$  €.

Ergibt es einen Unterschied, wenn man zunächst alle Mengen addiert, und alle Preise addiert und dann davon den Quotienten bildet? Hier kann dann der Begriff des gewichteten arithmetischen Mittels angesprochen werden.

Abb. 30: Erstellung der Listen und Berechnung der Mittelwerte

Menge	Preis	LPreis
1	45.23	6779/100
2	52.89	8193/100
3	35.89	5739/100
4	49.26	1507/20
5	22.45	3293/100
6	59.57	9227/100
7	42.49	6497/100

Calculation results:

- list2/list1 → LPreis: {1.49878, 1.54906, 1.59905, 1.52964, 1.46682}
- mean(LPreis) → k: 1.53162
- (sum(Preis)) / (sum(Menge)) → m: 1.53561

Abb. 31: Darstellung in Dezimalschreibweise

Menge	Preis	LPreis
1	45.23	67.79
2	52.89	81.93
3	35.89	57.39
4	49.26	75.35
5	22.45	32.93
6	59.57	92.27
7	42.49	64.97

list = main \ Preis

Der Kaufpreis, der sich aus dem mittleren Preis berechnet, der also theoretisch zu zahlen wäre, wird Schätzwert genannt. Der Schätzwert errechnet sich aus den Mengen und  $k$ . Die Liste heißt „schaetzp“ (s. Abb. 32, 34). Der Schätzwert liefert zu den einzelnen Kosten einen Unterschied, nennen wir ihn Abweichung. Ein „guter“ Mittelwert sollte eine kleine Abweichung haben (s. Abb. 33).

Abb. 32: Schätzwert

Preis	LPreis	schaetzp
1	67.79	1.49878
2	81.93	1.54906
3	57.39	1.59905
4	75.35	1.52964
5	32.93	1.46682
6	92.27	1.54893
7	64.97	1.52907

list = main \ schaetzp

Abb. 33: Abweichung des Schätzwertes zu den einzelnen Kosten

LPreis	schaetzp	abweich
1	1.49878	69.27527
2	1.54906	81.00749
3	1.59905	54.96992
4	1.52964	75.4477
5	1.46682	34.38492
6	1.54893	91.23873
7	1.52907	65.07862

list = main \ abweich

Beurteilt man die Ergebnisse, so sieht man, dass der Wert der Summe der Abweichungen der Einzelpreise vom Schätzwert sehr gering ist. Die Abweichung ist 1,22737, also nicht Null (s. Abb. 34). Wie steht es mit der Abweichung bezogen auf den Mittelwert  $m$ ? Der Fehler, die Abweichungssumme, ist Null. Die einzelnen Abweichungen schwanken erheblich. Der Mittelwert aus den Summen ist da schon ein sinnvolleres Maß, da die Abweichung nicht Null ist (s. Abb. 35).

The screenshot shows a calculator window with a table of data and a calculation of the sum of deviations. The table has three columns: 'LPreis', 'schaetzp', and 'abweich'. The data rows are numbered 4 to 10. Below the table, the 'list=' field contains 'main \ abweich'. The calculation area shows 'k \ Menge → schaetzp' with values {69.27526, 81.00749, 54.96991, 75.44770, 34}, 'Preis - schaetzp → abweich' with values {-1.48526, 0.92251, 2.42009, -0.09770, -1.4}, and 'sum (abweich)' resulting in 1.22737.

	LPreis	schaetzp	abweich
4	1.52964	75.4477	-0.0977
5	1.46682	34.38492	-1.45492
6	1.54893	91.23873	1.03128
7	1.52907	65.07862	-0.10862
8			
9			
10			

list= main \ abweich

k \ Menge → schaetzp  
 {69.27526, 81.00749, 54.96991, 75.44770, 34}

Preis - schaetzp → abweich  
 {-1.48526, 0.92251, 2.42009, -0.09770, -1.4}

sum (abweich)  
 1.22737

Abb. 34: Summe der Abweichungen der Einzelpreise vom Schätzwert

The screenshot shows a calculator window with a table of data and a calculation of the sum of deviations of the mean price from the estimated value. The table has three columns: 'LPreis', 'schaetzp', and 'abweich'. The data rows are numbered 1 to 7. Below the table, the 'list=' field contains 'main \ LPreis'. The calculation area shows 'm \ Menge → schaetzp' with values {69.45563, 81.21841, 55.11304, 75.64414, 34}, 'Preis - schaetzp → abweich' with values {-1.66563, 0.71159, 2.27696, -0.29414, -1.5}, and 'sum (abweich)' resulting in 0.00000.

	LPreis	schaetzp	abweich
1	1.49878	69.45563	-1.66563
2	1.54906	81.21841	0.7116
3	1.59905	55.11304	2.27696
4	1.52964	75.64414	-0.29414
5	1.46682	34.47444	-1.54444
6	1.54893	91.47628	0.79372
7	1.52907	65.24806	-0.27806

list= main \ LPreis

m \ Menge → schaetzp  
 {69.45563, 81.21841, 55.11304, 75.64414, 34}

Preis - schaetzp → abweich  
 {-1.66563, 0.71159, 2.27696, -0.29414, -1.5}

sum (abweich)  
 0.00000

Abb. 35: Summe der Abweichungen des Mittelwerts der Preise vom Schätzwert

**Fazit:**

Ziel der Untersuchung ist eine Schätzfunktion, die man auch lineare Regression nennt. Man kann sie so bestimmen, dass die Abweichungen der Einzelwerte klein sind. Aber die Abweichungen schwanken in einem großen Intervall. Die Abweichungen sind negative und positive Zahlen, damit lässt sich erklären, dass diese Abweichungssumme den Wert Null annimmt.

Möchte man diesen Mangel beseitigen, kann man die Beträge der Abweichungen addieren, dann sind die Einzelwerte nichtnegativ. Aber diese Summe soll möglichst klein sein, also minimal. Dazu stellt man umfangreiche Berechnungen an. Der unbekannte Wert für das Minimum muss durch systematisches Probieren, also schrittweises Einsetzen von einer Vielzahl von geschätzten Benzinpreisen gefunden werden. Man vergleicht die gefundenen Ergebnisse und sucht nach dem kleinsten Wert der Abweichungssumme. Das ist aber in der Klassenstufe 6 nur sehr eingeschränkt möglich.

Man kann auch den Zusammenhang Fahrstrecke und Benzinmenge untersuchen, dann kommt man auf den Begriff „Verbrauch“, genauer „durchschnittlicher Verbrauch“ an Benzin auf 100 km Fahrstrecke, so wie in Autowerbungen die Angaben erfolgen. Hier muss man beachten, dass die Werte der ersten Zeile nicht eingebracht werden können, da man die Betrachtung bei einer vollen Tankfüllung beginnt.

## Lösungshilfen zu Aufgabe 4: ICE von Hamburg nach München

Die Schülerinnen und Schüler erhalten das Blatt „Ihr Zugbegleiter“ eines ICE: Im Schülergespräch werden Fragen gefunden. Dann wird versucht, den Fragen nachzugehen.

Bahnhof	Zeitdauer in <i>h: min</i>	Strecke in <i>km</i>
Hamburg-Altona	00:00	0
Hamburg Dammtor	00:08	5
Hamburg Hbf.	00:06	1
Hamburg-Harburg	00:11	12
Lüneburg	00:21	38
Uelzen	00:25	35
Hannover	00:37	95
Göttingen	00:34	99
Kassel-Wilhelmshöhe	00:19	45
Fulda	00:30	90
Würzburg	00:31	93
Nürnberg	00:55	102
Ingolstadt	00:31	90
München Hbf.	00:49	81

Hier können alle bisher verwendeten Methoden eingesetzt werden. Es kann nach der Durchschnittsgeschwindigkeit zwischen Bahnhöfen gefragt werden. Dabei kann man die Zeiten des Stillstands berücksichtigen oder auch nicht. Es muss entschieden werden, ob man, wie üblich, Kilometer pro Stunde oder besser Kilometer pro Minute verwendet. Man kann die Daten kumulieren und so die zurückgelegte Strecke darstellen, als graphischen Fahrplan. Jeder Teilstrecke entspricht dann ein Kurvenzug.

Nach der Eingabe der Daten in Listen werden die Fahrzeiten addiert, dazu dient der Befehl *cuml*, der über *Aktion* → *Liste* → *Berechnung* zu finden ist (s. Abb. 36, 37). Ebenso addiert man alle Teilstrecken zu einer Gesamtstrecke „Gesamt“ (s. Abb. 38). Damit hat man nun auch die Möglichkeit, die Summe aller Daten zu bearbeiten.

	Stadt	Dauer	Strecke
1	HH_A	0	0
2	HH_D	8	5
3	HH_B	6	1
4	HH_H	11	12
5	LG	21	38
6	UE	25	35
7	H	37	95
8	GÖ	34	99
9	KS	19	45
10	FD	30	90
11	WÜ	31	93
12	N	55	102
13	IN	31	90
14	M	49	81

Abb. 36: Die Liste der Daten

	Dauer	Strecke	Fahrzeit
1	0	0	0
2	8	5	8
3	6	1	14
4	11	12	25
5	21	38	46
6	25	35	71

Abb. 37: Die Fahrzeit ohne Wartezeiten zusammengefasst als Fahrzeit

	Strecke	Fahrzeit	Gesamt
1	0	0	0
2	5	8	5
3	1	14	6
4	12	25	18
5	38	46	56
6	35	71	91

Abb. 38: Alle Listen der Daten

In Abbildung 39 sind die Einstellungen dargestellt, die vorgenommen werden müssen, um die Strecke über der Fahrzeit aufzutragen (s. Abb. 40).

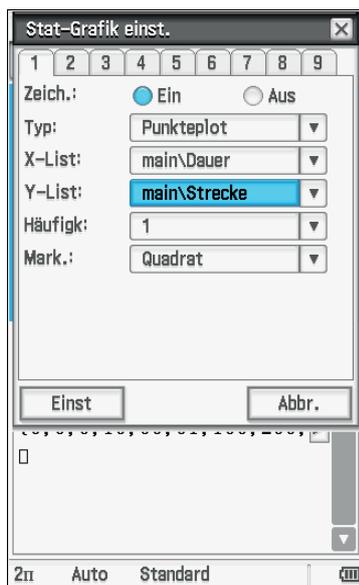


Abb. 39: Die Einstellungen für die Punktwolke Dauer und Strecke

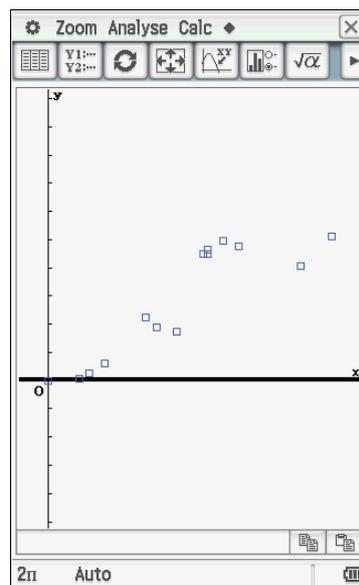


Abb. 40: Die Punktwolke Dauer/ Strecke

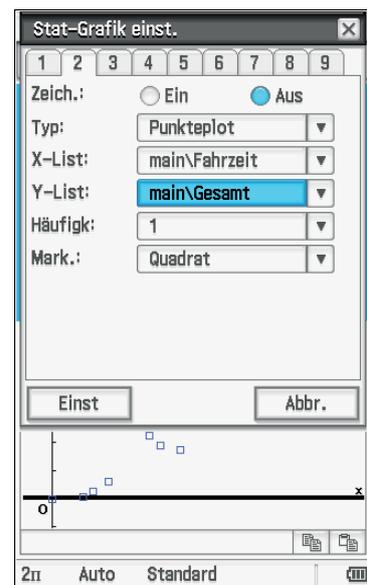


Abb. 41: Die Einstellung von Graph 2

Die Einstellungen für die Graphik der Gesamtstrecke und Fahrzeit sind in Abbildung 41 zu finden. In Abbildung 42 wird die Auswahl des zweiten Graphen gezeigt. Die zugehörige Darstellung der Punktwolke Fahrzeit über Gesamtstrecke zeigt Abbildung 43.

Die beiden Graphen zeigen deutlich, dass eine lineare Schätzfunktion einen Sinn ergibt (s. Abb. 40, 43). Der Proportionalitätsfaktor ist dann die Durchschnittsgeschwindigkeit des ICE.

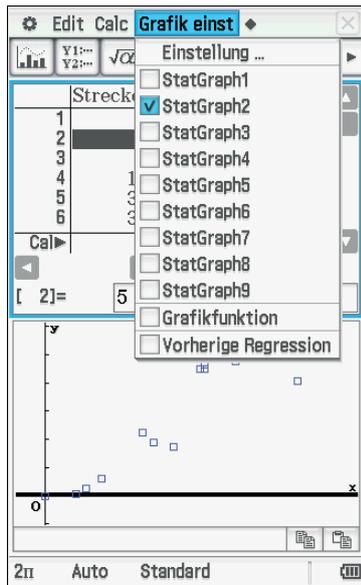


Abb. 42: Auswahl des Graphen 2

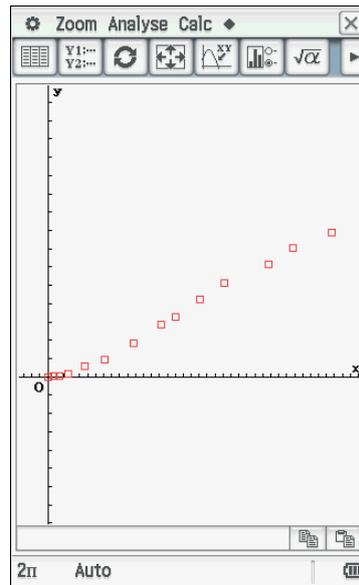


Abb. 43: Die Punktwolke Fahrzeit über Gesamtstrecke

Alle anderen Aufgaben werden analog zu den Aufgaben 1 und 2 bearbeitet.

### Anhang:

#### Theoretische Betrachtung der linearen Regression anhand des Einstiegsbeispiels:

„Für ein Modegeschäft wird in einem Zeitraum die Zahl  $x$  der vor dem Schaufenster stehenden Personen und die Zahl  $y$  der anschließend das Geschäft betretenden Kunden beobachtet.“

Zahl $x$ der die Auslagen ansehenden Personen	5	7	6	4	5	9
Zahl $y$ der den Laden betretenden Personen	3	5	4	2	4	6

Die gesuchte Gerade hat die Gleichung  $y = a + bx$ . Die Gerade geht durch den Punkt  $(\bar{x}, \bar{y})$ , den beiden Mittelwerten der eindimensionalen Datenreihen. Die Daten seien  $(x_i, y_i)$ .

Dann ist der Flächeninhalt  $A_1 = (y_1 - y)^2 = (y_1 - (a + bx_1))^2$ ;

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

Das definiert eine Funktion in zwei Variablen  $a$  und  $b$ :

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

Die Summe aller Flächeninhalte ist dann zu minimieren:

Das Minimum wird ermittelt über die beiden partiellen Ableitungen nach  $a$  und nach  $b$ .

$$\frac{\partial f(a,b)}{\partial a} = \sum_{i=1}^n 2 \cdot (y_i - (a + bx_i)) \cdot (-1),$$

$$\frac{\partial f(a,b)}{\partial b} = \sum_{i=1}^n 2 \cdot (y_i - (a + bx_i)) \cdot (-x_i).$$

Beide partielle Ableitungen werden Null gesetzt, das Gleichungssystem wird gelöst und liefert den stationären Punkt, der ein Minimum sein sollte.

$$\left\{ \begin{array}{l} \sum_{i=1}^n 2 \cdot (y_i - (a + bx_i)) \cdot (-1) = 0 \\ \sum_{i=1}^n 2 \cdot (y_i - (a + bx_i)) \cdot (-x_i) = 0 \end{array} \right\}$$

$$\left\{ \begin{array}{l} \sum_{i=1}^n y_i = a \cdot n + b \cdot \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i y_i = a \cdot \sum_{i=1}^n x_i + b \cdot \sum_{i=1}^n x_i^2 \end{array} \right\}$$

$$\left\{ \begin{array}{l} \frac{1}{n} \sum_{i=1}^n y_i = a + b \cdot \frac{1}{n} \cdot \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i y_i = a \cdot \sum_{i=1}^n x_i + b \cdot \sum_{i=1}^n x_i^2 \end{array} \right\}$$

Das führt zur Lösung für  $b$

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

Der Zähler ist bis auf einen Faktor die Kovarianz, der Nenner bis auf einen Faktor die Varianz von  $x$ .

Der  $y$ -Achsenabschnitt wird über die Geradengleichung mit einem bekannten Punkt gefunden.

$$\bar{y} = a + b \cdot \bar{x}$$

führt zu

$$a = \bar{y} - b \cdot \bar{x}.$$

Die Bedingung für ein Minimum kann man über die bekannte Beziehung der zweiten partiellen Ableitungen durch Einsetzen überprüfen. Damit hat man für die Schätzfunktion eine vollständige Lösung, die Steigung und den  $y$ -Achsenabschnitt.

Für beide Berechnungen verwendet man häufig praktische Tabellen als Arbeitshilfe.

Zurück zur Aufgabe:

1	2	3	4	5	6	7	8
$x_i$	$y_i$	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})^2$	$(x_i)^2$	$x_i \cdot y_i$	$(x_i - \bar{x}) \cdot (y_i - \bar{y})$
5	3	1	1	1	25	15	1
7	5	-1	-1	1	49	35	1
6	4	0	0	0	36	24	0
4	2	2	2	4	16	8	4
5	4	1	0	1	25	20	0
9	6	-3	-2	9	81	54	6
36	24			16	232	156	12
$\sum x_i$	$\sum y_i$			$\sum (x_i - \bar{x})^2$	$\sum (x_i)^2$	$\sum x_i \cdot y_i$	$\sum (x_i - \bar{x}) \cdot (y_i - \bar{y})$

Zur Berechnung von  $b$  benötigt man bei der ersten Darstellung die Summen der Spalten 7, 1, 2, 6, das Quadrat der Spaltensumme 1 sowie die Anzahl der Paare  $n$ .

$$b = \frac{n \cdot \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \cdot \sum_{i=1}^n y_i}{n \cdot \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2};$$

$$b = \frac{6 \cdot 156 - 36 \cdot 24}{6 \cdot 232 - 36^2} = \frac{3}{4};$$

Für die zweite Darstellung verwenden wir hingegen nur Spaltenwerte aus Spalte 8 und Spalte 5

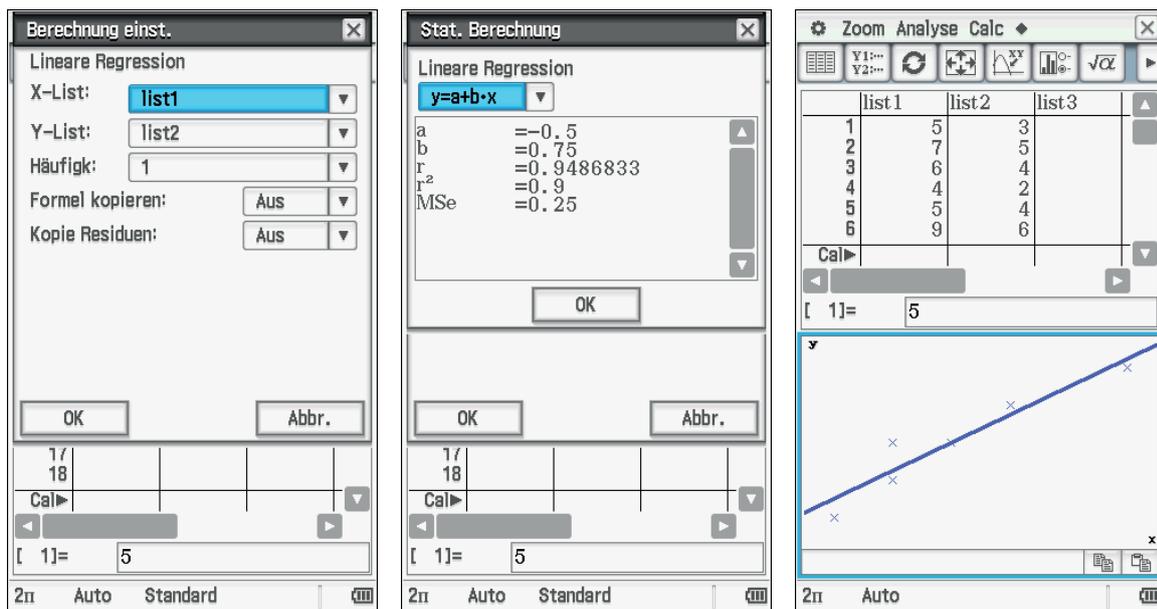
$$b = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$b = \frac{12}{16} = \frac{3}{4}$$

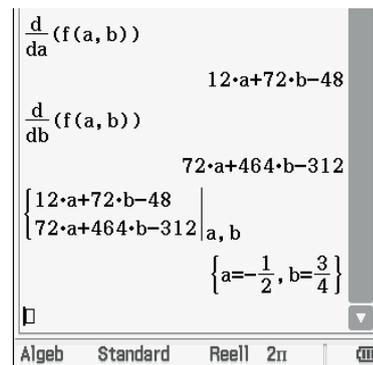
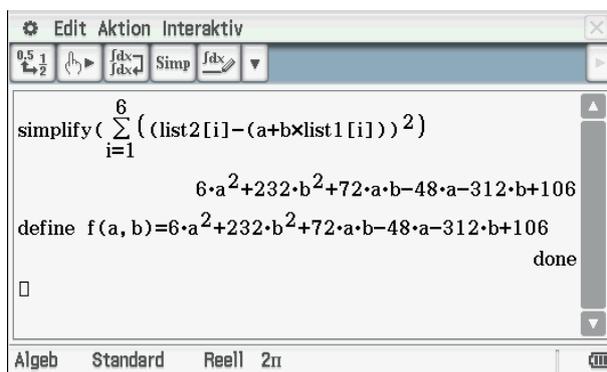
Der  $y$ -Achsenabschnitt ist  $a = -\frac{1}{2}$  und damit lautet die Gleichung der Schätzfunktion:  
 $y = -\frac{1}{2} + \frac{3}{4}x$ .

Damit ist mit einem Beispiel die Methode der linearen Regression beschrieben. Hier lässt sich zumindest der Steigungsfaktor  $\frac{3}{4}$  inhaltlich interpretieren: Wenn vier Personen vor dem Schaufenster stehen, so kamen drei davon in den Laden.

Das Erzeugen der Darstellung erfolgt mit dem ClassPad auf die folgende Weise:



Die Herleitung mit partiellen Ableitungen kann man im konkreten Fall mit dem ClassPad durchführen.



### Literatur:

Winter, H. (1996): Mathematikunterricht und Allgemeinbildung. Mitteilungen der Gesellschaft für Didaktik der Mathematik Nr. 61, 37–46.