

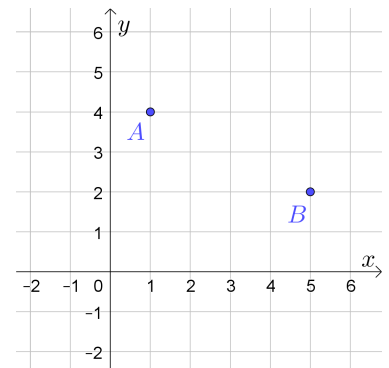
Rechts unten sind zwei Punkte A und B mit verschiedenen x -Koordinaten dargestellt. Es gibt genau eine **lineare Funktion** f , deren Graph durch A und B verläuft.

- 1) Zeichne den Funktionsgraphen rechts ein.
- 2) Ermittle eine Funktionsgleichung von f .

$f(x) =$

- 3) Vom Punkt P ist die x -Koordinate bekannt. Berechne mithilfe von f durch **lineare Interpolation** die y -Koordinate von P .

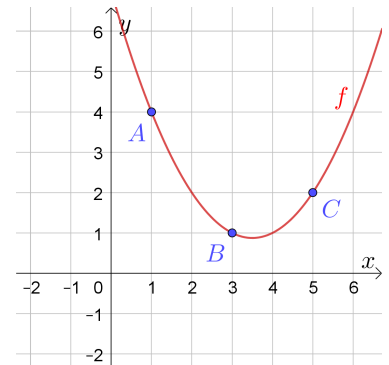
$P = (4,2 |$)



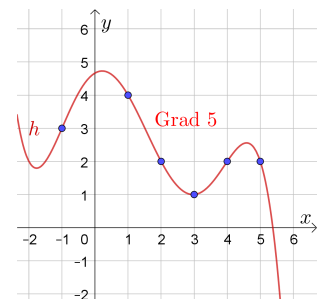
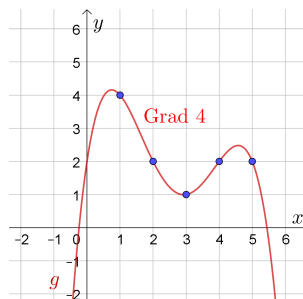
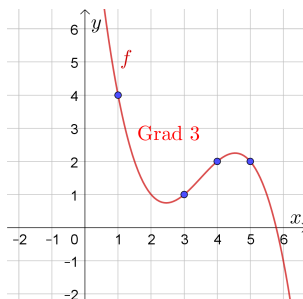
Rechts unten sind drei Punkte A , B und C mit paarweise verschiedenen x -Koordinaten dargestellt, die *nicht* auf einer Geraden liegen. Es gibt genau eine **quadratische Funktion** $f(x) = a \cdot x^2 + b \cdot x + c$, deren Graph durch A , B und C verläuft.

- 1) Stelle ein Gleichungssystem zur Berechnung von a , b und c auf.
- 2) Ermittle a , b und c mit **Technologieeinsatz**.
- 3) Ermittle jene x -Koordinate in $[1; 5]$ so, dass der Punkt P am Funktionsgraphen liegt.

$P = ($ $| 2,3)$



Allgemein können wir durch $n + 1$ Punkte an verschiedenen Stellen eine **Polynomfunktion** legen. Dafür reicht eine Polynomfunktion, deren Grad höchstens ist.



Ermittle eine Funktionsgleichung der rechts oben dargestellten Polynomfunktion h vom Grad 5. In GeoGebra kannst du dafür auch den Befehl **TrendPoly**(<Liste von Punkten>, <Grad des Polynoms>) verwenden.

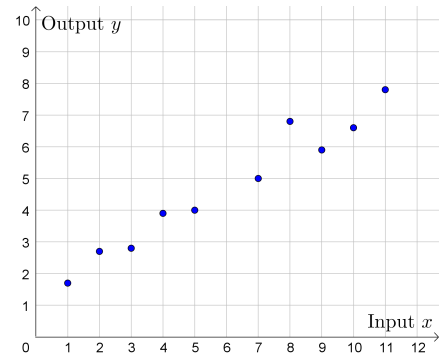
$h(x) =$

Modellierung  **MmF**

Wir führen eine Messreihe durch.
Abhängig vom Input x erhalten wir einen Output y als Messwert.

Die Messergebnisse sind rechts dargestellt.

Mit welchem Output y rechnest du ungefähr,
wenn der Input $x = 6$ ist?



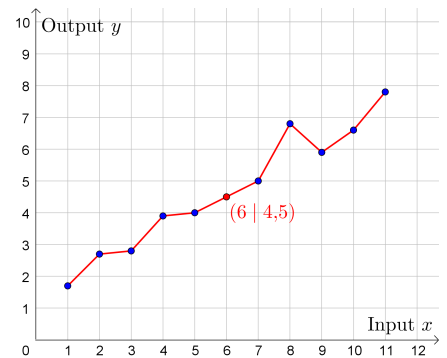
Stückweise lineare Interpolation  **MmF**

Rechts arbeiten wir mit **stückweise linearer Interpolation**.

Wir verbinden dabei von links nach rechts benachbarte Messpunkte durch eine Strecke.

So erhalten wir den Graphen einer stückweise linearen Funktion.

Beim Input $x = 6$ sagen wir mit stückweise linearer Interpolation den Output $y = 4,5$ voraus.



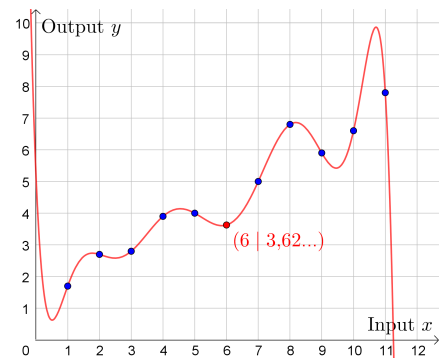
Overfitting  **MmF**

Rechts arbeiten wir mit Polynominterpolation.

Wir legen also eine Polynomfunktion *exakt* durch die Messpunkte.

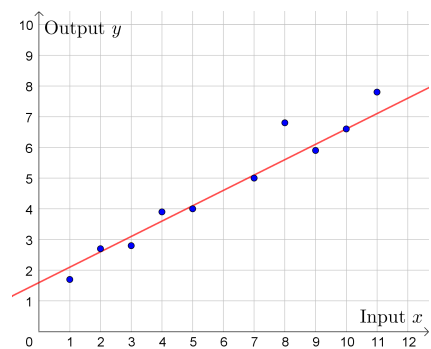
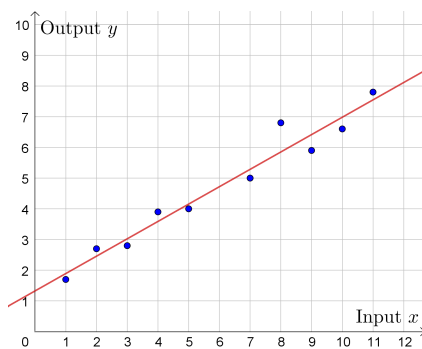
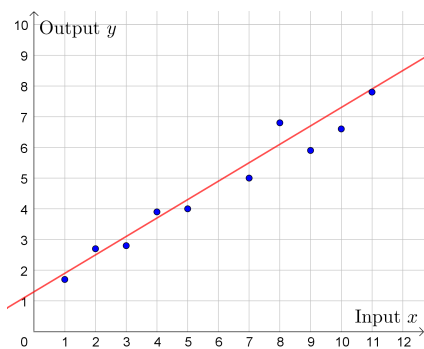
Beim Input $x = 6$ sagen wir mit Polynominterpolation den Output $y = 3,62\dots$ voraus.

An den Stellen $x = 0,1$ und $x = 10,5$ macht diese Polynomfunktion aber vermutlich *keine* guten Voraussagen.



Lineare Regression  **MmF**

Bei **linearer Regression** legen wir die *bestmögliche* lineare Funktion durch die Punktwolke.
Welche der 3 dargestellten linearen Funktionen passt aus deiner Sicht am besten?



Wir sind auf der Suche nach *einer* Gerade, die den Zusammenhang zwischen Input und Output unter Berücksichtigung aller Messpunkte *am besten* beschreibt.

Rechts haben wir eine Gerade $y = k \cdot x + d$ durch die Punktwolke gelegt.

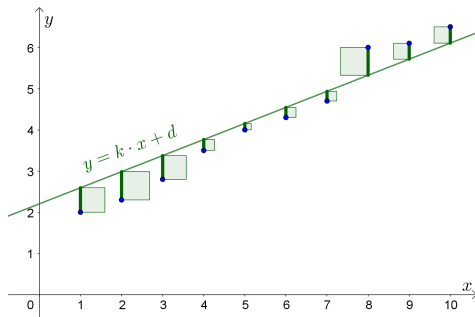
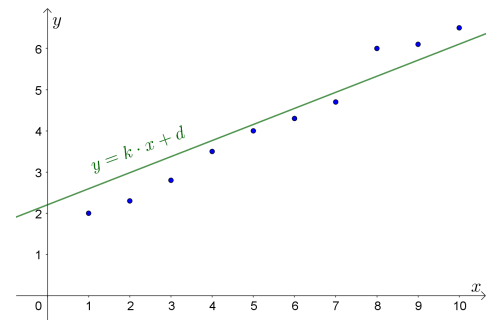
Die eingezeichnete Gerade ist gut, aber nicht perfekt.

Sollen wir die Steigung k größer oder kleiner wählen?

Sollen wir die Gerade vertikal verschieben?

Dazu müssen wir festlegen, was wir mit *am besten* meinen.

Wie sollen wir den „Fehler“ quantifizieren?



Um die sogenannte **Fehlerquadratsumme** zu ermitteln, berechnen wir zuerst die vertikalen Abstände der Messpunkte von der Gerade:

0,595	0,685	0,575	0,265	0,155
0,245	0,235	0,675	0,385	0,395

Die **Fehlerquadratsumme** F ist die Summe der Quadrate dieser Abstände.

Links siehst du die Fehlerquadrate veranschaulicht.

Für die Fehlerquadratsumme dieser Gerade gilt: $F = 0,595^2 + 0,685^2 + \dots + 0,395^2 = 2,12325$

Die Fehlerquadratsumme hängt von der Gerade ab, die wir durch die Punktwolke legen. $F(k; d)$

Je kleiner die Fehlerquadratsumme, desto besser ist die Gerade an die Punktwolke angepasst.

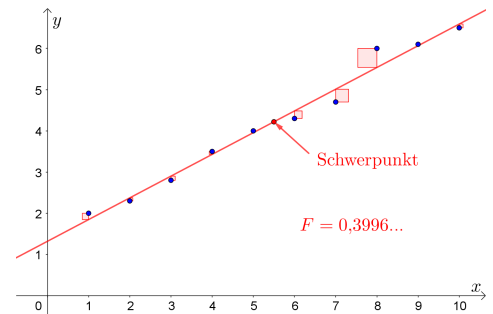
Die *bestmögliche* Gerade ist jene mit der kleinsten Fehlerquadratsumme.

Man kann die beiden folgenden Aussagen beweisen:

- 1) Zu jeder Punktwolke mit n Punkten gibt es *genau eine* bestmögliche Gerade ($n \geq 2$).
- 2) Diese bestmögliche Gerade verläuft durch den **Schwerpunkt** $(\bar{x} | \bar{y})$ der Punktwolke.

\bar{x} ist das **arithmetische Mittel** der x -Koordinaten aller Messpunkte.

\bar{y} ist das arithmetische Mittel der y -Koordinaten aller Messpunkte.



Diese Gerade durch die Punktwolke mit der kleinsten Fehlerquadratsumme heißt **Regressionsgerade**.

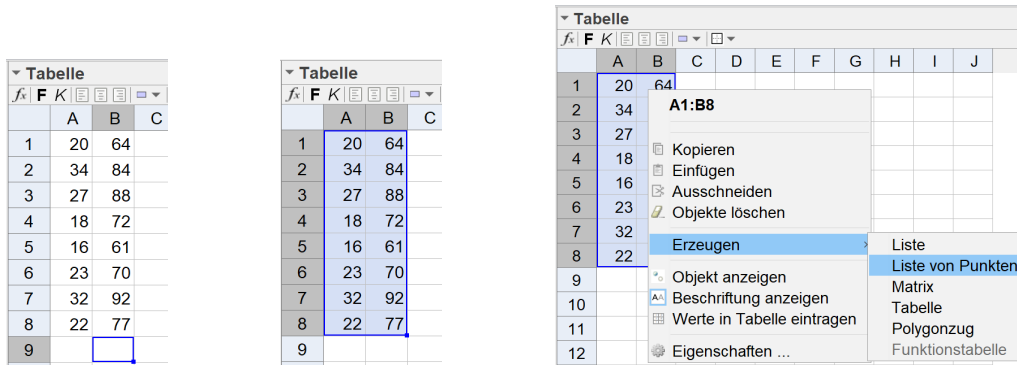
Die zugehörige lineare Funktion heißt **lineare Ausgleichsfunktion**.

Von einer Schülergruppe wurden die jeweilige Lernzeit (in Minuten) und die erreichte Punktezahl bei einer Leistungsüberprüfung notiert:

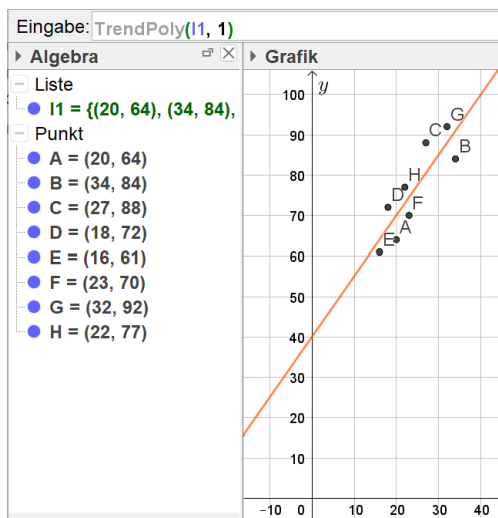
Lernzeit in Minuten	20	34	27	18	16	23	32	22
erreichte Punktezahl	64	84	88	72	61	70	92	77

- 1) Ermitteln Sie die Gleichung der zugehörigen Regressionsgeraden. (Die erreichte Punktezahl soll in Abhängigkeit von der Lernzeit beschrieben werden.)
- 2) Interpretieren Sie die Steigung der Regressionsgeraden in diesem Sachzusammenhang.
- 3) Berechnen Sie mithilfe dieses Modells, welche Punktezahl man erwarten kann, wenn man 30 Minuten lernt.

- i) Tabellenansicht öffnen (Ansicht → Tabelle)
- ii) Daten eingeben (Unabhängige Größe x in Spalte A, Abhängige Größe y in Spalte B)
- iii) Daten markieren → Rechtsklick → Erzeugen → Liste von Punkten



iv) Lineare Ausgleichsfunktion ermitteln: **TrendPoly**(<Liste von Punkten>, <Grad des Polynoms>)



1) $f(x) =$

2) Interpretation der Steigung:

3) Punktezahl bei 30 Minuten Lernzeit:

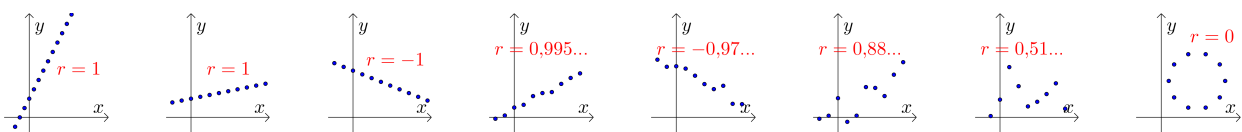
Korrelationskoeffizient nach Pearson 


$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ sind n Wertepaare. \bar{x} bzw. \bar{y} ist das arithmetische Mittel der x_i bzw. y_i . Der **empirische Korrelationskoeffizient** r misst den Grad des linearen Zusammenhangs.

Er hat folgende Eigenschaften:

- $-1 \leq r \leq 1$
- $r = 1 \iff$ Alle Punkte liegen auf *einer* Gerade mit positiver Steigung.
- $r = -1 \iff$ Alle Punkte liegen auf *einer* Gerade mit negativer Steigung.
- Ein Korrelationskoeffizient nahe bei 1 bzw. -1 bedeutet einen starken positiven bzw. negativen linearen Zusammenhang. Sprechweise: „Die Daten korrelieren.“

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



 **KorrelationsKoeffizient**(<Liste von Punkten>)

Der Gesamtumsatz von Fairtrade-Produkten in Österreich ist in den letzten Jahren deutlich gestiegen:

Jahr	2007	2008	2009	2010	2011	2012	2013
jährlicher Gesamtumsatz in Millionen (Mio.) Euro	53	65	72	87	100	107	130

Quelle: http://www.fairtrade.at/fileadmin/AT/Materialien/2013_FAIRTRADE_Inside_Zahlen_Fakten.pdf [05.09.2016].

Die nachstehende Abbildung zeigt diese Gesamtumsatzentwicklung.

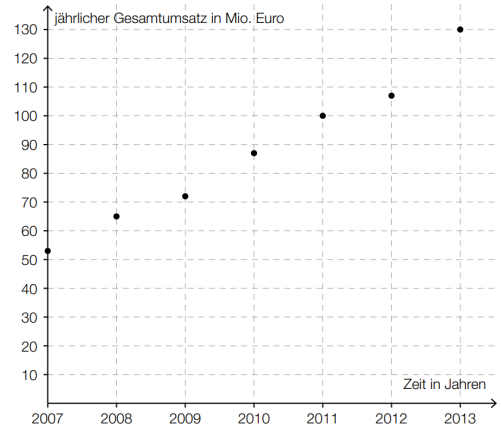
Der jährliche Gesamtumsatz soll in Abhängigkeit von der Zeit beschrieben werden.

- 1) Ermitteln Sie mithilfe der gegebenen Daten eine Gleichung der zugehörigen linearen Regressionsfunktion. Wählen Sie $t = 0$ für das Jahr 2007.

$f(t) =$

$t \dots$ Zeit in Jahren ($t = 0$ entspricht dem Jahr 2007)
 $f(t) \dots$ jährlicher Gesamtumsatz zur Zeit t in Mio. Euro

- 2) Zeichnen Sie den Graphen der Regressionsfunktion im nebenstehenden Koordinatensystem ein.
- 3) Beurteilen Sie mithilfe des Korrelationskoeffizienten, ob die lineare Regressionsfunktion ein geeignetes Modell zur Beschreibung der Gesamtumsatzentwicklung ist.
- 4) Berechnen Sie anhand dieses Modells den zu erwartenden jährlichen Gesamtumsatz im Jahr 2020.



Korrelation \neq Kausalität

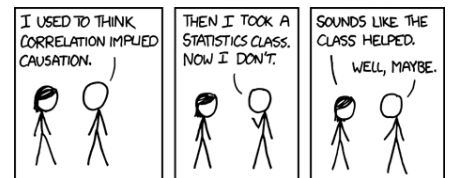


Eine Korrelation bedeutet *nicht* automatisch einen kausalen Zusammenhang (Ursache \rightarrow Wirkung).

Zum Beispiel gibt es folgende Korrelationen:

- a) „Je mehr Feuerwehrpersonen zum Einsatz kommen, desto größer ist der Schaden.“
- b) „Je höher die Verkaufszahlen von Eis in Kalifornien, desto größer ist die Anzahl der Haiattaken.“

Was könnte jeweils ein Grund für die Korrelation sein? Nimm dazu Stellung.



Quelle: <https://xkcd.com/552>

Weitere Ausgleichsfunktionen



Quadratische Ausgleichsfunktion:	TrendPoly (<Liste von Punkten>, 2)	$f(x) = a \cdot x^2 + b \cdot x + c$
Kubische Ausgleichsfunktion:	TrendPoly (<Liste von Punkten>, 3)	$f(x) = a \cdot x^3 + b \cdot x^2 + c \cdot x + d$
Exponentielle Ausgleichsfunktion:	TrendExp (<Liste von Punkten>)	$f(x) = c \cdot e^{\lambda \cdot x}$
	TrendExp2 (<Liste von Punkten>)	$f(x) = c \cdot a^x$
Logarithmische Ausgleichsfunktion:	TrendLog (<Liste von Punkten>)	$f(x) = a + b \cdot \ln(x)$

