

1. Einführung

Häufig steht man im Labor vor der Aufgabe einen Zusammenhang zwischen einem Messsignal und einer Gehaltsgröße herzuleiten. Diese Zusammenhänge sind dem Betrachter von Diagrammen auch schon mit dem Auge leicht nachvollziehbar.

Ziel der *Regressionsanalyse* ist es, eine mathematische Modellfunktion $f(x)$, herauszufinden, die den Zusammenhang zwischen den beiden Größen beschreibt. Bei der Ermittlung der Modellfunktion wird diese so angepasst, dass sie, etwas vereinfacht gesprochen, die Abstände zu den Messpunkten (Residuen) minimiert werden (Streng genommen wird die Summe der quadratischen Abweichungen wird minimiert).

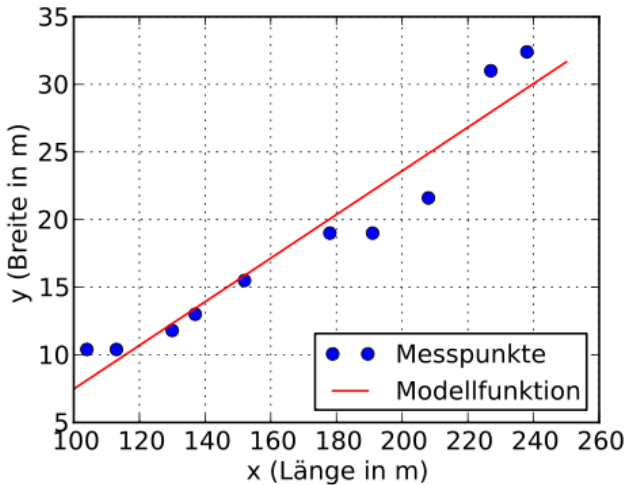


Abb. 1: xy-Diagramm (Streudiagramm) von Längen und Breiten von 10 zufällig ausgewählten Kriegsschiffen mit eingezeichneter linearer Modellfunktion. Quelle:

wikipedia.de

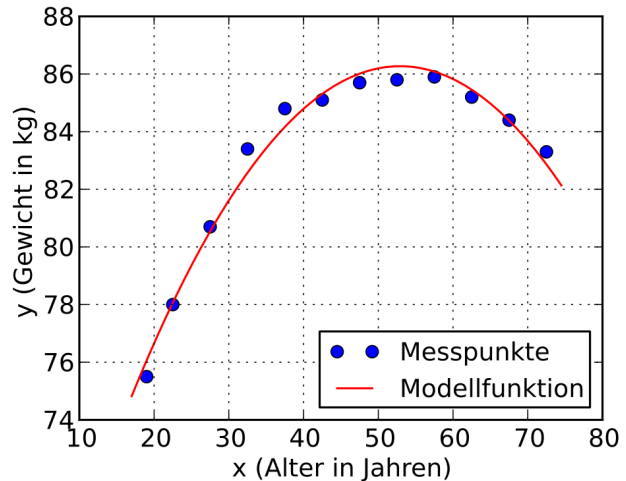


Abb. 2: xy-Diagramm: Durchschnittliches Gewicht von Männern nach Alter mit eingezeichneter parabelförmiger Modellfunktion Quelle: wikipedia.de

Mithilfe der Modellfunktionen kann man dann **interpolieren**, d.h. man kann innerhalb des modellierten Datenbereichs anhand der einen Größe, die andere abschätzen. Man erkennt beispielsweise anhand Abb. 1, dass ein Kriegsschiff mit einer Breite von 20 Metern ungefähr eine Länge von Metern haben müsste. Männliche Personen im Alter von 30 Jahren wiegen im Schnitt ca..... Kilo wiegen. Statt mithilfe eines Diagramms auf einem Millimeterpapier oder einem Ausdruck eine graphische Interpolation mit dem Auge durchzuführen („das Lot fällen“), kommt man zu präziseren Ergebnissen, wenn man den interpolierten Wert mittels der mathematischen Gleichung der Modellfunktion berechnet.

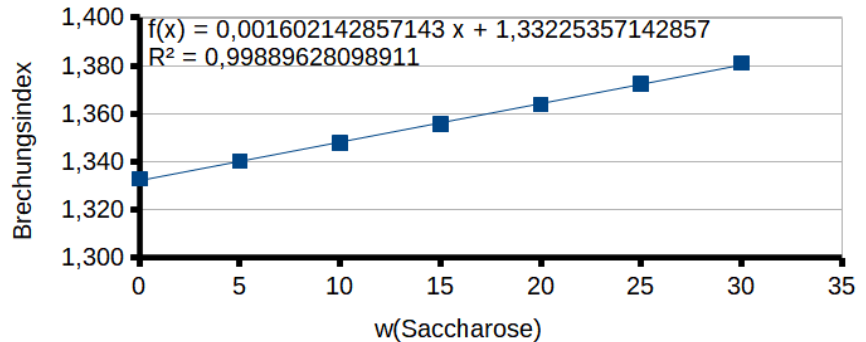
Mithilfe verschiedener Programme, lässt sich die am besten passende Funktionsgleichung bestimmen. Man gibt dem Programm die allgemeine Form der Funktionsgleichung der Modellfunktion vor, z.B. bei linearer Modellfunktion, $f(x) = a \cdot x + b$ und Startwerte für die Parameter (z.B. bei einer Geraden: a und b vor. Das Programm verändert jetzt ausgehend von den Startwerten die Parameter a und b in kleinen Schritten. Es berechnet dann das Ergebnisse von $f(x)$ und vergleicht mit den tatsächlichen Messwerten. Je kleiner diese Unterschiede ausfallen, desto besser sind die gerade benutzten Parameter a und b. Nach vielen Schritten hat es die optimalen Parameter gefunden. Solche Methoden, die letzten durch schrittweise Variation und Ausprobieren die optimalen Parameter ermitteln werden **numerische Methoden** genannt.

2. Lineare Kalibrierung

Die Anzeige der meisten Analyseinstrumente gibt meistens ein Signal aus, dass linear vom Gehalt abhängt.

1. Die Kalibrierung eines Refraktometers liefert folgende Ergebnisse:

| w(Sac.) in % | Brechungsindex |
|--------------|----------------|
| 0 | 1,3330 |
| 5 | 1,3403 |
| 10 | 1,3478 |
| 15 | 1,3557 |
| 20 | 1,3638 |
| 25 | 1,3723 |
| 30 | 1,3811 |



a) Berechnen Sie den Gehalt einer Probe, wenn der Brechungsindex 1,3511 beträgt.

b) Prüfen Sie die Gültigkeit folgender Behauptung: *Verdoppelt man den Gehalt, so verdoppelt sich das Messsignal.*

Viertelt man den Gehalt, so viertelt sich auch das Messsignal.

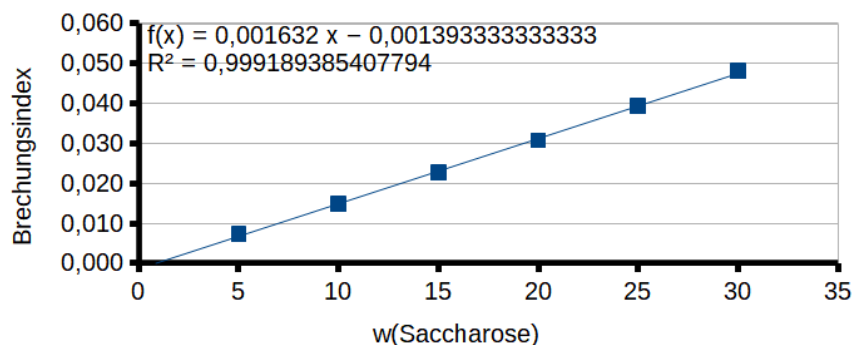
Zur einfacheren Interpretation zieht man es häufig vor, einen linearen Zusammenhang zu erhalten, bei dem ein Gehalt von Null, ein Messsignal von Null erzeugt. Dann sind Messsignal und Gehalt zueinander **proportional**: Liegt z.B. doppelter Gehalt vor, so ist auch das Ergebnis ungefähr doppelt so hoch. Viertelt man den Gehalt, so liefert das Gerät ein Ergebnis, dass nur ca. 1 Viertel des ursprünglichen Wertes beträgt. Dies führt auch zu einer Kalibriergerade, die keinen oder nur einen kleinen y-Achsenabschnitt besitzt. Hierfür kann man meistens so verfahren, dass man eine Lösung mit einem Gehalt von Null (**Blindprobe**) misst und das Messergebnis (**Leerwert**, **Blindwert**, **Reagenzienleerwert**) von den anderen Messwerten abzieht. Die allermeisten Geräte lassen es heute zu, dass der Blindwert auf Tastendruck auf Null gestellt wird. So entfällt das manuelle Abziehen des Wertes von den anderen Messergebnissen.

Es gilt zu beachten, dass die Blindprobe alle anderen Stoffe (Puffer, Lösungsmittel, Reagenzien) in der Quantität enthält, wie sie auch in den anderen Kalibrierproben vorkommen, außer eben den Analyt. Ist bekannt, dass bestimmte Stoffe den Messwert nicht beeinflussen, z.B. Puffersubstanzen, verzichtet man manchmal der Einfachheit halber, auf deren Anwesenheit in der Blindprobe und setzt nur das Lösungsmittel selbst ein.

Das **Einstellen des Reagenzienleerwertes auf Null** wird auch häufig „**Messen gegen die Referenz**“ oder „**Referenz auf Null stellen**“ o.ä. bezeichnet. Im Biolaborjargon spricht man auch von „**blanken**“, weil eine Blindprobe ohne Gehalt im Englischen *blank solution* heißt.

2. a) Geben Sie bei dem Beispiel von oben bleibend (Aufgabe 1), die Messwerte nach Einstellung des Leerwertes auf Null, in untenstehender Tabelle an.

| w(Sac.) in % | Brechungsindex |
|--------------|----------------|
| 5 | |
| 10 | |
| 15 | |
| 20 | |
| 25 | |
| 30 | |



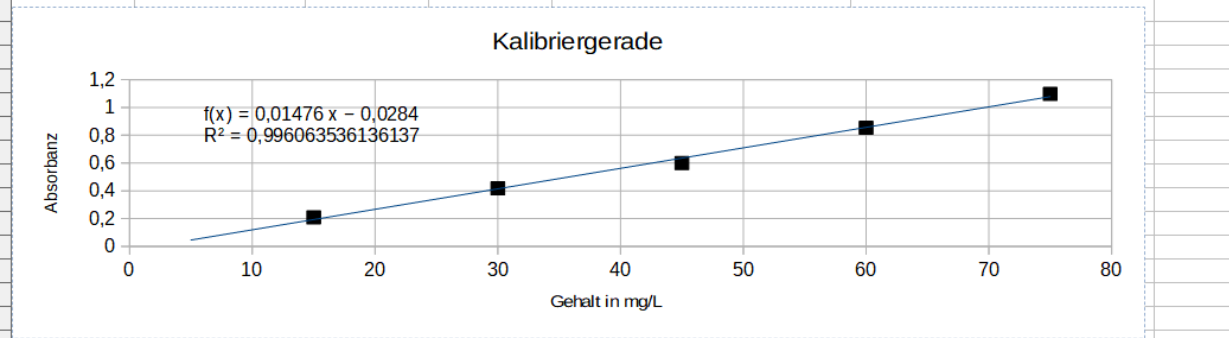
b) Die Probe zeigt jetzt ein Brechungsindex von 0,0181. Zeigen Sie, dass dasselbe Ergebnis wie oben resultiert.

2.1 Kalibriergeraden erstellen mit LibreOffice Calc

Erklär-YouTube-Film (auf englisch): <https://www.youtube.com/watch?v=7CgxEjXVEsM>

Beispiel einer mit einem Tabellenkalkulationsprogramm erstellten Kalibriergerade (incl. Interpolation):

| A | B | C | D | E | F | G |
|----------|----------------|--------------|---|---------------------------------|---------------------------------|---|
| | Gehalt in mg/L | Absorbanz | | | Durch Funktion angezeigter Wert | |
| Kalibr 1 | 15 | 0,209 | | Steigung | 0,01476 | |
| Kalibr 2 | 30 | 0,418 | | y-Achsenabschnitt | -0,0284 | |
| Kalibr 3 | 45 | 0,6 | | Erwartete Absorbanz für Probe 1 | 0,330268 | |
| Kalibr 4 | 60 | 0,854 | | Erwarteter Gehalt für Probe 2 | 39,0749066075371 | |
| Kalibr 5 | 75 | 1,098 | | | | |
| Probe1 | 24,3 | siehe rechts | | | | |
| Probe2 | siehe rechts | 0,548 | | | | |



2.2 Vorgehensweise zur Erstellung mit LibreOffice Calc

Zuerst erstellt man eine Tabelle mit den Messpunkten. Spalte links: Gehalt (z.B. mg/L oder % o.ä.). Spalte rechts daneben: Messwert (z.B. Absorbanz, Dicht o.ä.). Anschließend in Menüleiste durchklicken: → Einfügen → Diagramm erstellen → XY-Streudiagramm → Datenbereich angegeben → Diagramm fertig stellen lassen.

Durch Doppelklick auf das Diagramm mit anschließendem Rechtsklick auf einen Datenpunkt kann man sich eine **Trendlinie einfügen** (= Kalibriergerade) und auch die Geradengleichung anzeigen lassen. Weitere Möglichkeiten betreffen das Verlängern der Gerade über den ersten und letzten Messpunkt hinaus. Lassen Sie auch das **Bestimmtheitsmaß (R^2)** anzeigen. Es ist ein Maß, wie gut der lineare Zusammenhang ist. Idealwert bei absolut linearem Zusammenhang: 1,0000.

Weitere nützliche Funktionen in **Libreoffice Calc** sind **STEIGUNG (SLOPE*)** und **ACHSENABSCHNITT (INTERCEPT*)**, mit denen man sich diese Werte der Geradengleichung in Zellen anzeigen lassen kann. So können Sie z.B. zum Weiterrechnen benutzt werden. Sehr wichtig ist auch die Funktionen **ERWARTUNG (FORECAST*)**, **TREND** oder auch **PROGNOSE** (je nach installierter Version), mit der man Werte interpolieren kann. Das ist das, was man händisch auf der Kalibriergeraden durch *Lot fällen* auf die x-Achse oder auf die y-Achse ermitteln kann. Hat die Probe1 z.B. den Gehalt 24,3 mg/L (**siehe Abb.**), so kann man die dazugehörige Absorbanz auf ca. 0,330268 (**siehe Abb.**) interpolieren lassen. Vertauscht man die x-Datenreihe und die y-Datenreihe, so kann man für eine gegebene Absorbanz der Probe 2 den Gehalt interpolieren lassen (**siehe Bsp. in Abb. ca. 39,07 mg/L**). Dies entspricht der typischen Erfordernissen einer Versuchsauswertung! Alle drei Funktionen greifen direkt auf die Daten zu, eine Darstellung eines Diagramms ist nicht erforderlich.

Aufgabe: Erstellen Sie ein entsprechendes Kalibrierdiagramm mit Ihren Daten aus dem Praktikum (Alternative auf Lehreranweisung hin: Daten vom Beispiel oben nutzen).

- Berechnen Sie den Gehalt der Probe mit dem Taschenrechner und der Geradengleichung der Näherungsfunktion.
- Nutzen Sie zur Berechnung des Gehalts der Probe die oben vorgestellten Funktionen.

*Je nach installierter Sprache werden die deutschen oder die in Klammern angegeben englischen Funktionsnamen angezeigt.

2.3 Kalibriergerade erstellen und Standardabweichung berechnen mit Taschenrechner und App

Die meisten wissenschaftlichen Taschenrechner modernerer Bauart können Kalibriergeraden (d.h. die Steigung und den y-Achsenabschnitt) anhand von Wertepaaren ermitteln. Das ist ein großer Vorteil, weil auch ohne Computer im Labor die Ergebnisse ermittelt werden können! Für häufige Taschenrechner gibt es hierzu auch Anleitungen im Netz.

Erklär-YouTube-Filme: einfach youtube und als Suchbegriffe „linear(e) regression“ + TR-Marke (Modell) eingeben.

Gleiches gilt auch für häufige statistische Lage- und Streumaße (z.B. arithmetisches Mittel, Standardabweichung)

Erklär-YouTube-Filme: einfach youtube und als Suchbegriffe „Standardabweichung“ oder „standard deviation“ + TR-Marke

Mini-Anleitungen für gängige Casio- und Texas Instruments Taschenrechner

2.3.1 Casio, z.B. fx-85, fx-991 etc.

Standardabweichung

- Mit **MODE** in den STAT-Modus wechseln. Anschließend 1-VAR auswählen (das sind die Kennzahlen auf einer einfachen Werteliste beruhen, nicht Wertepaare).
- In der nun erscheinenden Tabelle die Einzelwerte eingeben. Jede Eingabe mit **=**-Taste quittieren.
- **Beispiel:** 12,2 15,6 14,3 13,8 13,9
- Am Ende mit der **AC**-Taste („All Clear“) die Dateneingabe beenden.
- Mit **SHIFT** **1** wieder in das Statistikmenu wechseln und VAR auswählen. Hier kann man sich den Mittelwert (\bar{x}), und die Standardabweichung einer Stichprobe (s oder σ_{n-1}) anzeigen lassen. Achtung: nicht σ wählen, weil dies die Standardabweichung ausgehend von einer Grundgesamtheit ist.

Ergebnisse für das Beispiel von oben: $\bar{x} = 13,96$ **s = 1,217784874**

Regressionsgerade

- Mit **MODE** in den STAT-Modus wechseln. Anschließend A+BX auswählen, also den Ausdruck einer Geradengleichung.
- In der nun erscheinenden Tabelle die X/Y-Wertepaare eingeben:
Beispiel: 25/0,245 50/0,480 75/0,738 100/0,985 125/1,230
- Am Ende mit der **AC**-Taste („All Clear“) die Dateneingabe beenden.
- Mit **SHIFT** **1** in (Zweitfunktion über der Taste 1) in das Statistikmenu wechseln und *Reg* für Regressionsanalyse auswählen. Im Untermenü kann man sich A (Achsenabschnitt) und B (Steigung) der Geradengleichung anzeigen lassen. Quadriert man den Wert von r, erhält man das Bestimmtheitsmaß R^2 anzeigen lassen.

Ergebnisse für Bsp von oben: **A = $-6,9 \cdot 10^{-3}$ und B = $9,9 \cdot 10^{-3}$, also $y = 9,9 \cdot 10^{-3} \cdot x - 6,9 \cdot 10^{-3}$ $R^2 = 0,9998454274$**

2.3.2 Texas Instruments, z.B. TI-30X Plus

Standardabweichung

- Mit **data** erst die Werteliste eingeben. Alle Werte in einer Spalte (z.B. in L1) eingeben. Jede Eingabe mit **enter** quittieren.
Beispiel: 12,2 15,6 14,3 13,8 13,9
- Anschließend **clear** drücken. Der Taschenrechner hat die Werteliste nach wie vor gespeichert.
- Mit **2nd** **data** (Zweitfunktion über der **data**-Taste) in das Statistikmenu wechseln und dort *1-Var Stats* (das sind die Kennzahlen auf einer einfachen Werteliste beruhen, nicht Wertepaare) auswählen.
- Im sich öffnenden Menü überprüfen ob die betreffende Daten-Spalte auch vorausgewählt wurde (z.B. L1).
- FRQ-Zeile so lassen wie sie ist. Auf **CALC** scrollen und **enter** drücken.

Ergebnisse für das Beispiel von oben: \bar{x} = 13,96 s_x = 1,217784874 Es werden auch andere statistische Größen angezeigt, wie Median, Quadratsumme, Minimalwert etc. Achtung: Der Wert bei σ entspricht der Standardabweichung ausgehend von einer Grundgesamtheit, und ist für uns nicht relevant. Wir brauchen immer Standardabweichung aus einer Stichprobe.

Regressionsgerade

- Mit **data** erst die Werteliste eingeben. Alle x-Werte in einer Spalte (z.B. in L1) eingeben, die dazugehörigen y-Werte in der Nachbarspalte (L2). Jede Eingabe mit **enter** quittieren.

Beispiel: 25/0,245 50/0,480 75/0,738 100/0,985 125/1,230

- Anschließend **clear** drücken. Der Taschenrechner hat die Werteliste nach wie vor gespeichert.
- Mit **2nd data** (Zweitfunktion über der **data**-Taste) in das Statistikmenu wechseln und dort *LinReg* aufrufen.
- Im sich öffnenden Menu überprüfen ob die betreffende Daten-Spalten auch vorausgewählt wurde
- FRQ-Zeile so lassen wie sie ist. Auf **CALC** scrollen und **enter** drücken.

Ergebnisse für Bsp von oben: A = -0,0099 und B = 0,0069; also $y = 0,0099^3 \cdot x - 0,0069$ $R^2 = 0,999845427$

3. In einer Wasserprobe wurde fotometrisch der Nitratgehalt untersucht. Dazu wurden 20 mL der Probe auf 100 mL verdünnt. Zu 10 mL dieser Verdünnung (oder Kalibrierlösungen mit den angegebenen Gehalt) wurden 15 mL Reagenzien zugesetzt und auf 50 mL mit Wasser aufgefüllt. Nach kurzer Wartezeit wurde die Absorbanzen gemessen. Bestimmen Sie den Gehalt an Nitrat in den Proben...

- ausschließlich mit dem Taschenrechner. Geben Sie auch die Standardabweichung und den Variationskoeffizienten an.
- [wenn im U behandelt]:** nachdem Sie vorher mit dem Verfahren von GRUBBS auf Ausreißer geprüft und ggf. entsprechende Messwerte eliminiert haben. Geben Sie auch die Standardabweichung und den Variationskoeffizienten an.
- ... mithilfe eines Tabellenkalkulationsprogramms. **[wenn im U behandelt]:** Identifizieren Sie zuvor Ausreißer nach GRUBBS indem Sie mit Formeln/Befehlen in einer Zelle den Prüfwert berechnen lassen.

Kalibrierdaten (nach Nullwertanpassung mit einem Reagenzienleerwert):

| | | | | | |
|---------------------------------|-------|-------|-------|-------|-------|
| $\beta(\text{NO}_3^-)$ in mg/L: | 15 | 30 | 45 | 60 | |
| Absorbanz: | 0,238 | 0,560 | 0,836 | 1,142 | |
| Absorbanzen der Proben: | 0,538 | 0,544 | 0,428 | 0,484 | 0,584 |

Bestimmen Sie den Gehalt aller Proben, die Standardabweichung und den Variationskoeffizient.

Lösung: Kalibriergerade: $f(x) = 0,01992 x - 0,053$. Proben: 148,4 mg/L; 149,9 mg/L; 120,8 mg/L; 134,8 mg/L; 159,9 mg/L
Mittelwert: 142,76 mg/L Standardabweichung: 15,2 mg/L Variationskoeffizient: 10,65%