Regression und Residuum

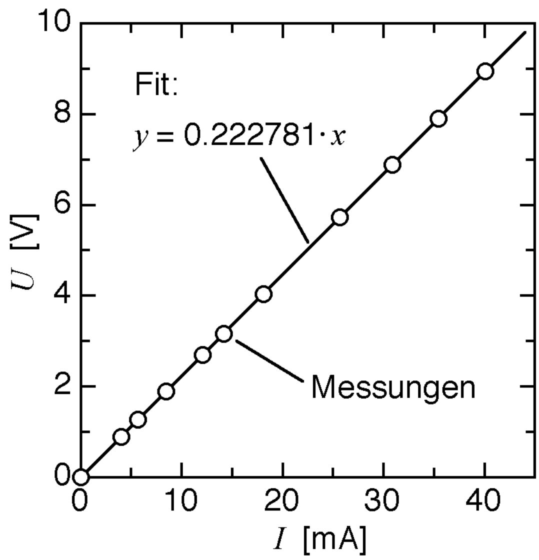
Lie.

Ein Residuum (lat. Rest, Plural Residuen) ist das was übrig bleibt, wenn vom Messwert ein berechneter Wert abgezogen wird. Mit Residuen wird die Güte eines Fits beurteilt. Ein Fit, engl. Anpassung, ist eine Regressionsfunktion.

U [V]	I [mA]
0.000	0.000
0.893	4.011
1.268	5.69
1.883	8.45
2.687	12.08
3.158	14.16
4.04	18.14
5.72	25.68
6.88	30.90
7.90	35.45
8.94	40.12

Tabelle 1: Strom I durch ein Widerstandselement, wenn es an ein geregeltes Netzgerät mit Gleichspannung U angeschlossen wird. Zwei Digitalmultimeter lieferten die Daten. Die Fehlerschranken waren 0.01 V resp. 0.01 mA. Das Widerstandselement hatte den Nennwert 220 Ω (±10 %) und Maximalleistung 5 W. (Lie. April 06)

Das Widerstandselement sollte das Ohmsche Gesetz erfüllen, d.h. der Strom sollte proportional zur Spannung anwachsen. Wir wollen den Widerstandswert durch Fit einer Proportionalität bestimmen.



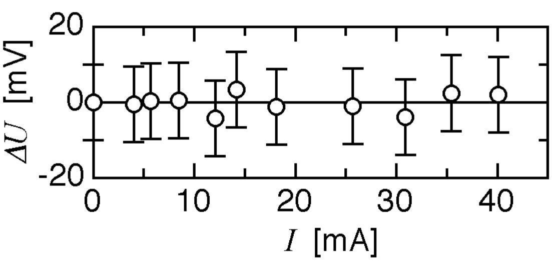


Abb. 1: Oben: Strom-Spannungs-Messwerte mit einer Regressionsfunktion ($x \mathrel{\hat{=}}$ Strom I in mA, $y \mathrel{\hat{=}}$ Spannung U in Volt).
Unten: Residuen ΔU in mV (!) mit Fehlerbalken (die Fehlerbalken für den Strom sind zu klein zum einzeichnen).

Der Vergleich der Regressionsfunktion mit $U = RI$ lässt einen direkt den Widerstandswert ablesen:
$R = 0.222781\ \text{V/mA} = 222.781\ \Omega$

Die Residuen sind gleichmässig um Null herum verteilt. Alle Residuen sind innerhalb der Fehlerschranken (10 mV) Null. Der Fit passt also gut zu den Daten.

Fehlerbetrachtung: Erhöht man die grösste Spannung in der Messreihe um die Fehlerschranke so resultiert die Proportionalitätskonstante 0.222857. Es sind also nicht mehr als fünf wesentlichen Ziffern gesichert: R = 222.78 Ω.

Herleitung der Regressionsformel für eine Proportionalität

Gesucht ist jene Proportionalität $y = ax$, die am besten zu einer Menge von Wertepaaren (x_i, y_i) ; $i = 1,..,n$ passt. Was "am besten" bedeutet, ist nicht ganz eindeutig. Carl Friedrich Gauss (1777-1855, dt. Mathematiker und Physiker) entwickelte die Regression nach der Methode der kleinsten Quadrate, weil sich diese einfach rechnen lässt. Nach dieser ist die Anpassung optimal, wenn die Summen der Quadrate der Differenzen von Messwert und Funktionswert minimal wird, d.h. wenn die Summe der Quadrate der Residuen minimal wird.
Mathematisch sieht diese Bedingung so aus:

$$\sum_{i=1}^{n} \left(y_i - y(x_i)\right)^2 \overset{hier}{=} \sum_{i=1}^{n} \left(y_i - ax_i\right)^2 = Minimum$$

Den mittleren Term kann man ganz einfach umformen:

$$\sum_{i=1}^{n} \left(y_i - ax_i\right)^2 = \sum_{i=1}^{n} \left(y_i^2 - 2ax_iy_i + a^2x_i^2\right) = \sum_{i=1}^{n} y_i^2 + \sum_{i=1}^{n} \left(-2ax_iy_i\right) + \sum_{i=1}^{n} a^2x_i^2$$

Der letzte Term ist eine quadratische Funktion von a. Dies sieht man sofort, wenn wir die konstanten Ausdrücke durch die Grossbuchstaben A, B, C ersetzten:

$$\left(\sum_{i=1}^{n} x_i^2\right)a^2 + \left(\sum_{i=1}^{n} -2x_iy_i\right)a + \left(\sum_{i=1}^{n} y_i^2\right) = Aa^2 + Ba + C$$

Ein quadratischer Term $Aa^2 + Ba + C$ hat das Minimum (Scheitel der Parabel) bei

$$a = \frac{-B}{2A} = \frac{-\left(\sum_{i=1}^{n} -2x_iy_i\right)}{2\left(\sum_{i=1}^{n} x_i^2\right)} \qquad \Leftrightarrow \qquad a = \frac{\sum_{i=1}^{n} x_iy_i}{\sum_{i=1}^{n} x_i^2}$$

Man kann zeigen, dass bei einer solchermassen bestimmten Regressionsfunktion die Summe der Residuen verschwindet, d.h. die Summe der positiven Residuen ist betragsmässig gleich der Summe der negativen Residuen.
Die Methode der kleinsten Quadrate hat noch weitere, statistisch günstige Eigenschaften. Deshalb wird sie heute praktisch überall eingesetzt. Jeder anständige technisch/wissenschaftliche Taschenrechner hat sie einprogrammiert.

Die Berechnung der Fehlerschranken der Regressionskonstanten ist kein Mittelschulstoff mehr.