

Lineare Regression einer Geraden mit R-GNU

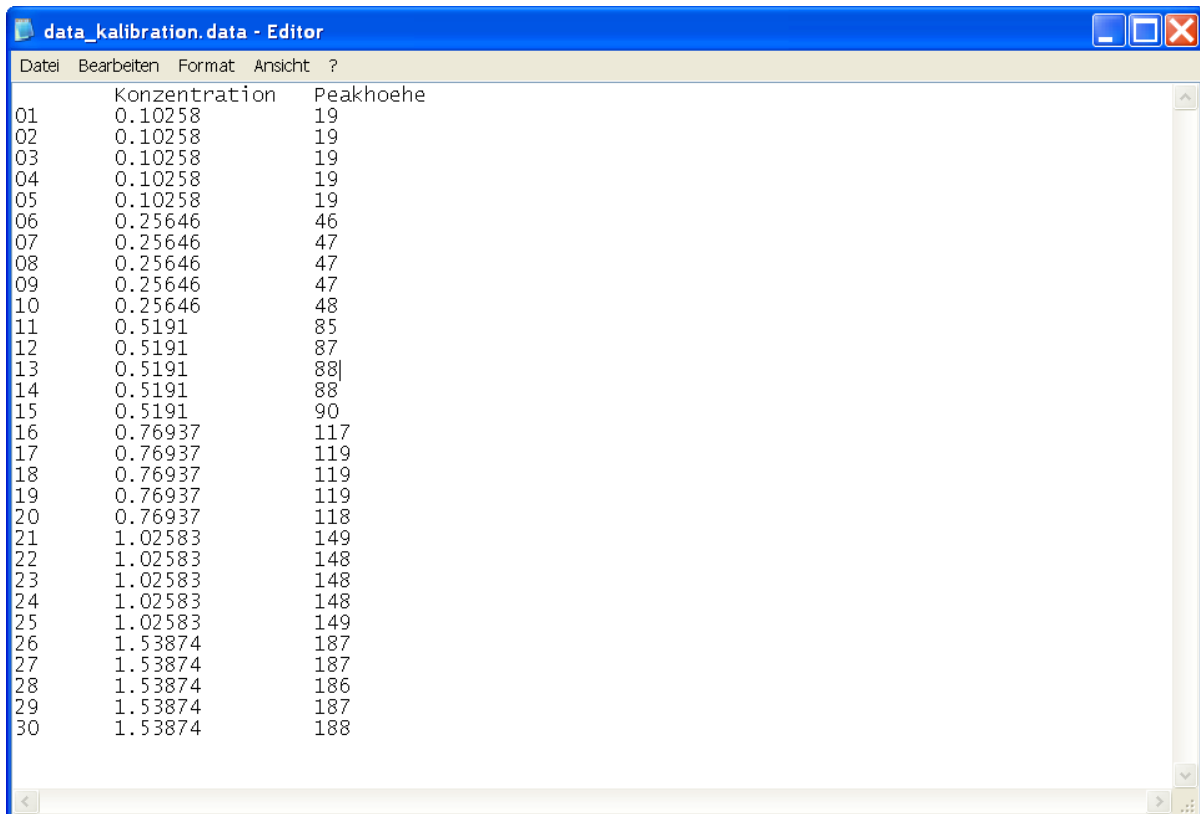
11.4.2007 Joël Gubler

Was ist GNU R?¹

GNU R ist eine freie Statistik-Software die unter GNU-Lizenz steht. R lehnt sich an die in den Bell Laboratories entwickelte Sprache S an. R konnte sich auf allen wesentlichen Bereichen der Angewandten Statistik etablieren.

Datenaufbereitung

Als erster Schritt müssen die Messdaten so aufbereitet werden, dass sie von R gelesen werden können. Die Konzentrationen der Standards und die zugehörigen Signale werden einem Editor als Kolonne aufgetragen und ein Laufindex hinzugefügt. Wichtig ist es die Kolonnen zu benennen (bei Fig.1 Konzentration und Peakhoehe). Die erstellte Datei wird als *data_kalibration.data* gespeichert.



	Konzentration	Peakhoehe
01	0.10258	19
02	0.10258	19
03	0.10258	19
04	0.10258	19
05	0.10258	19
06	0.25646	46
07	0.25646	47
08	0.25646	47
09	0.25646	47
10	0.25646	48
11	0.5191	85
12	0.5191	87
13	0.5191	88
14	0.5191	88
15	0.5191	90
16	0.76937	117
17	0.76937	119
18	0.76937	119
19	0.76937	119
20	0.76937	118
21	1.02583	149
22	1.02583	148
23	1.02583	148
24	1.02583	148
25	1.02583	149
26	1.53874	187
27	1.53874	187
28	1.53874	186
29	1.53874	187
30	1.53874	188

Figure 1 Eingabe der Kalibrationsdaten

Einstellen der Umgebung

Dem Programm R muss angegeben werden, wo es das nun erstellte Datenfile finden kann. Beim Rechtsklick auf eine Verknüpfung der Startdatei von R, kann bei den Eigenschaften bei *Ausführen in:* der Pfad der Ausführung festgelegt werden. Der Ordner der als Pfad in Anführungs- und Schlusszeichen angegeben wird, soll die Datei *data_kalibration.data* enthalten.

¹ Artikel GNU_R aus der Wikipedia am 29.4.2006

http://de.wikipedia.org/w/index.php?title=GNU_R&oldid=16125946

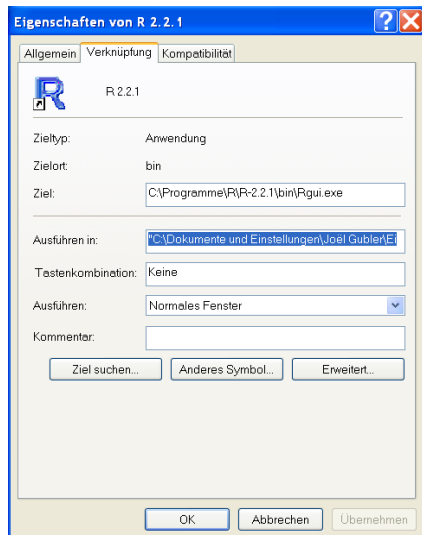


Figure 2 Ausführort der Verknüpfung ändern

Eingaben

Folgende Anweisungen werden in die Kommandozeile von R nacheinander eingegeben:

```
kalib<-read.table("data_kalibration.data")
```

```
plot(Peakhoehe~Konzentration,data=kalib,xlab="Konzentration in mmol/l",ylab="Peakhöhe in mm")
```

```
title("Kalibration Phosphorsäurebestimmung ( 26.4.2006)")
```

```
abline(lm(Peakhoehe~Konzentration,data=kalib))
```

```
summary(lm(Peakhoehe~Konzentration,data=kalib))
```

Dabei entsteht Fig.3.

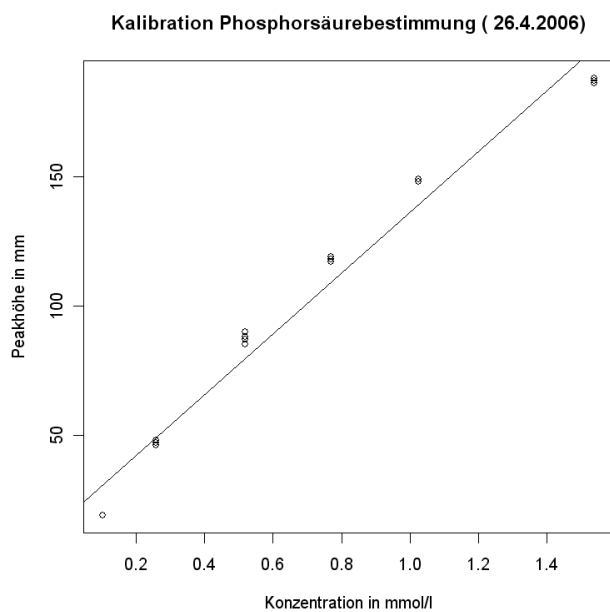


Figure 3 Plot Kalibration

- Die erste Eingabe importiert die Daten von *data_kalibration.data* als ein Objekt vom Typ *data frame* und benennt dieses Objekt kalib.
- Die Funktion `plot()`, trägt die Kolonne namens *Peakhoehe* gegen diejenige namens *Konzentration* aus dem Objekt kalib auf, und benennt die Achsen der Grafik mit *Konzentration in mmol/l* und *Peakhöhe in mm*.
- Mit `title()` wird der Grafik ein Titel gegeben.
- `abline(lm())` zeichnet die Regressionsgerade ein.
- `summary(lm())` gibt detaillierte Informationen zur Regression aus.

Ausgabe nach Abrufung des `summary()`:

Call:

```
lm(formula = Peakhoehe ~ Konzentration, data = kalib)
```

Residuals:

```
   Min      1Q  Median      3Q      Max
-13.280 -11.993  2.088  8.822 10.200
```

Coefficients:

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   18.972      3.147   6.028 1.70e-06 ***
Konzentration 117.179      3.693 31.730 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 9.774 on 28 degrees of freedom
Multiple R-Squared:  0.9729,    Adjusted R-squared:  0.972
F-statistic: 1007 on 1 and 28 DF,  p-value: < 2.2e-16
```

Erklärungen zum Output

`lm()` generiert Objekt vom Typ Liste, welches Informationen zur Regression enthält. Mit `summary()` werden die Listeneinträge zusammengefasst dargestellt.

Die Grundlage für die lineare Regression ist die allgemeine lineare Gleichung:

$$y = a + bx \quad \text{mit Achsenabschnitt } a \text{ und Steigung } b$$

Zum Berechnen einer Regressionsfunktion, variiert der Computer die Parameter der Modellfunktion (z.B. $y=a+bx$ mit Parameter a und b). Die Parameter werden so verändert, bis die Summe der Quadrate aus den Unterschieden der Messwerte zu den berechneten Werten, minimal ist. Der Unterschied heisst auf Englisch *residual* und die berechneten Werte *fitted values*.

Residuals: Hier steht Information zu den residuals

Coefficients: Information über die Grösse und Güte der Parameter

- Intercept(Achsenabschnitt): 18.972 mit 3.147 Std. Error (Standardabweichung)
- Konzentration(Empfindlichkeit):117.179 mit 3.693 Std. Error (Standardabweichung)

R-squared(R^2): 0.972 ist Bestimmtheitsmass

Achsen mit griechischen Buchstaben beschriften

Die Beschriftung von Achsen erfolgt mit `xlab` (resp. `ylab`) das entweder wie oben in der Funktion `plot()` oder danach in `title()` aufgerufen wird.

```
title(xlab=expression(paste(„vorne“, mu, „hinten“)))
```

- `title()` fügt einem Plot Beschriftungen hinzu
- Was `xlab` (resp. `ylab`) zugewiesen wird muss vom Typ `character` oder `expression` sein. Mit `expression` lassen sich mathematische Ausdrücke notieren, und somit auch griechische Buchstaben und weiteres darstellen (z.B. hoch- und tiefgestellter Text).
- Weil nicht nur mathematische „Formeln“ dargestellt werden sollen, sondern gleichzeitig noch Text, wird `paste()` verwendet, das einzelne R-Objekte verknüpft. Weil die Funktion `paste()` innerhalb von `expression()` steht, werden mathematische Ausdrücke verstanden und von `paste()` mit zu einer Zeichenkette zusammengesetzt.

Literatur

- Maindonald J. H.: *Using R for Data Analysis and Graphics*; Centre for Bioinformation Science, Australian National University, 2004, pp 28, 37-39
- Venables W. N., Smith D. M.: *An Introduction to R*; R manual, 2005