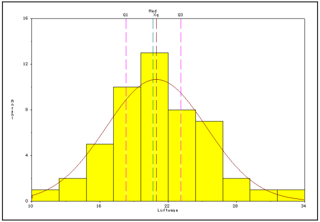
12. Datenanalyse: Inhalt

Histogramme

Aus Messwerten derselben Größe kann man eine Häufigkeitsverteilung in Form eines Histogramms erstellen, das die Messwerte in Intervalle (“Bins”) einträgt.



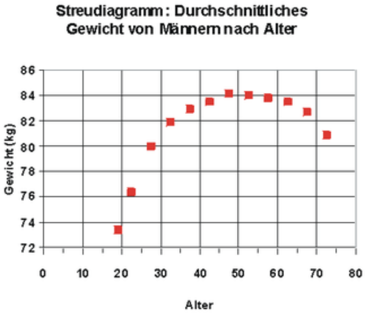
Zu beachten:

- Ein Histogramm enthält weniger Information als die Messwerte selbst.
- Der Mittelwert hängt von der Bineinteilung ab, insbesondere bei wenigen Messwerten.
- Man sollte weder zu wenige Bins, noch zu wenige Messwerte pro Bin haben.
- Die Bins sollten die gleiche Breite haben, damit die Höhe statt der Fläche abgelesen werden kann.

Streudiagramm

Ein Streudiagramm ist die grafische Darstellung von beobachteten Wertepaaren zweier Größen.

Normalerweise wird die Messung mit der größeren Unsicherheit auf der y-Achse dargestellt.



Zu beachten:

- Ist die Unsicherheit der Messwerte bekannt, so sollte sie auch dargestellt werden.
- Die Dimension der Punkte sollte so gewählt werden, dass die Unsicherheit, falls dargestellt, sichtbar ist.

Anpassung von Messdaten mit einer Funktion

Bei einer Messung erwartet man, dass die Werte eine bestimmte Häufigkeitsverteilung (z.B. eine Gaußverteilung) haben, oder dass eine gewisse Abhängigkeit (z.B. linear) zwei Größen verbindet.

Eine Anpassungsrechnung ("Fit") ist eine Optimierungsmethode, um die unbekannten Parameter der Funktion so zu bestimmen, so dass die Messwerte am besten durch die Funktion approximiert werden.

Es muss ein Maß für die Abweichung zwischen Messwerten und Funktion definiert werden, das minimiert werden kann.

Die Minimierung erfolgt meistens numerisch und kann sehr zeitaufwändig sein.

Die Funktion muss sorgfältig gewählt werden. Zu wenige freie Parameter geben eine schlechte Übereinstimmung, zu viele sind anfällig für statistische Fluktuationen.

Methode der kleinsten Quadrate

Seien (x_i, y_i) n gemessene Wertepaare, an die die Funktion $f(x_i; \boldsymbol{\theta})$ durch die zu bestimmenden Parameter $\boldsymbol{\theta} = \{\theta_1, \ldots, \theta_m\}$ angepasst werden soll. Die Unsicherheiten auf y_i seien σ_i (die Unsicherheiten auf x_i seien vernachlässigbar) und V sei die Kovarianzmatrix.

Das Maß für die Abweichung zwischen Punkten und Funktion ist

Chi-Quadrat

$$\chi^2(\boldsymbol{\theta}) = \sum_{i,j=1}^{n} \big(y_i - f(x_i; \boldsymbol{\theta})\big) \left(V^{-1}\right)_{ij} \big(y_j - f(x_j; \boldsymbol{\theta})\big) \tag{194}$$

Für unkorrelierte Messwerte ist die Kovarianzmatrix diagonal und es gilt

$$\chi^2(\boldsymbol{\theta}) = \sum_{i=1}^{n} \frac{\big(y_i - f(x_i; \boldsymbol{\theta})\big)^2}{\sigma_i^2} \tag{195}$$

Methode der kleinsten Quadrate

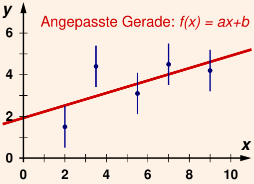
Beispiel: Anpassung einer Geraden an Messwerte

An 5 Messwerte (s. Tabelle) soll die Gerade

$$f(x) = a\,x + b$$

angepasst werden. Die freien Parameter sind $\theta = \{a, b\}$, also Steigung a und Offset b.

i	x_i	y_i
1	2,0	$1{,}5 \pm 1{,}0$
2	3,5	$4{,}4 \pm 1{,}0$
3	5,5	$3{,}1 \pm 1{,}0$
4	7,0	$4{,}5 \pm 1{,}0$
5	9,0	$4{,}2 \pm 1{,}0$



Zur Anpassung der freien Parameter a und b wird das χ^2 gebildet:

$$\chi^2(a,b) = \sum_{i=1}^{5} \frac{\big(y_i - f(x_i;a,b)\big)^2}{\sigma_i^2} = \sum_{i=1}^{5} \frac{\big(y_i - (a\,x_i + b)\big)^2}{\sigma_i^2}$$

$$= \sum_{i=1}^{5} \frac{a^2 x_i^2 + 2ab\,x_i + b^2 - 2a\,x_i y_i - 2b\,y_i + y_i^2}{\sigma_i^2}.$$

Methode der kleinsten Quadrate

Die χ^2-Funktion muss bezüglich a und b minimiert werden, d.h. die entsprechenden Ableitungen müssen 0 sein (s. Seite 78):

$$\frac{\partial \chi^2}{\partial a} = \sum_{i=1}^{5} \frac{2ax_i^2 + 2bx_i - 2x_i y_i}{\sigma_i^2} = 2a \sum_{i=1}^{5} \frac{x_i^2}{\sigma_i^2} + 2b \sum_{i=1}^{5} \frac{x_i}{\sigma_i^2} - 2 \sum_{i=1}^{5} \frac{x_i y_i}{\sigma_i^2} \stackrel{!}{=} 0$$

$$\frac{\partial \chi^2}{\partial b} = \sum_{i=1}^{5} \frac{2ax_i + 2b - 2y_i}{\sigma_i^2} = 2a \sum_{i=1}^{5} \frac{x_i}{\sigma_i^2} + 2b \sum_{i=1}^{5} \frac{1}{\sigma_i^2} - 2 \sum_{i=1}^{5} \frac{y_i}{\sigma_i^2} \stackrel{!}{=} 0$$

Die Lösung dieses linearen Gleichungssystems mit den Unbekannten a und b ist unschön und man nimmt am Besten einen Computer.

In unserem Fall sind aber alle Fehler σ_i gleich, damit erhält man einfacher:

$$\boldsymbol{a} = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{x^2} - \overline{x}^2} = \mathbf{0{,}299} \quad \text{und} \quad \boldsymbol{b} = \frac{\overline{x^2} \cdot \overline{y} - \overline{xy} \cdot \overline{x}}{\overline{x^2} - \overline{x}^2} = \mathbf{1{,}927}.$$

Die Gerade mit diesen Parametern ist im Diagramm auf der vorigen Seite.

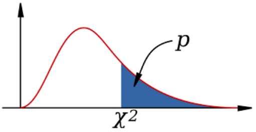
Methode der kleinsten Quadrate

Güte der Anpassungsrechnung/des Fits

Meistens ist es wichtig zu wissen, ob die angepasste Funktion die Daten gut beschreibt. Ein Maß dafür ist der Wert $\chi^2(\hat{\theta})$ der angepassten Parameter $\hat{\theta}$. Je niedriger er ist, desto besser stimmen die Datenpunkte innerhalb ihrer Fehler mit der Fitfunktion überein.

Häufig verwendet man den **reduzierten χ^2-Wert $\chi^2/n_{\text{d.o.f.}}$**.
Die Zahl der Freiheitsgrade $n_{\text{d.o.f.}}$ (*degrees of freedom*) ist die Zahl der Datenpunkte minus die Zahl der angepassten Parameter. Für einen guten Fit sollte $\chi^2/n_{\text{d.o.f.}} \approx 1$ sein.

Ein weiteres Maß für die Güte des Fits ist der ***p*-Wert**, der sich aus χ^2 und $n_{\text{d.o.f.}}$ berechnet und die Wahrscheinlichkeit angibt, dass man bei erneuter Durchführung des Experiments einen größeren χ^2-Wert bekommen würde.



Methode der kleinsten Quadrate

Beispiel: Güte des Geradenfits

Der χ^2-Wert der Geradenanpassung des vorigen Beispiels (Seiten 186/187) ist mit $\sigma_1 = \cdots = \sigma_5 = \sigma$

$$\chi^2(a,b) = \sum_{i=1}^{5} \frac{\left(y_i - (a\,x_i + b)\right)^2}{\sigma_i^2} = \frac{1}{\sigma^2} \sum_{i=1}^{5} \left(y_i - (a\,x_i + b)\right)^2$$

$$= \frac{1}{\sigma^2} \sum_{i=1}^{5} \left(a^2 x_i^2 + 2ab\,x_i + b^2 - 2a\,x_i y_i - 2b\,y_i + y_i^2\right).$$

$$= \frac{5}{\sigma^2} \left(a^2\,\overline{x^2} + 2ab\,\overline{x} + b^2 - 2a\,\overline{xy} - 2b\,\overline{y} + \overline{y^2}\right) = \mathbf{3{,}73}.$$

Die Zahl der Freiheitsgrade ist: $n_{\text{d.o.f.}} = 5$ Messwerte $-$ 2 Parameter $= 3$.

$\Longrightarrow \quad \chi^2 / n_{\text{d.o.f.}} = \frac{3{,}71}{3} = 1{,}24 \quad \Longrightarrow \quad p\text{-Wert} = 29{,}4\,\% \quad \Longrightarrow$ Fit ok. ✔