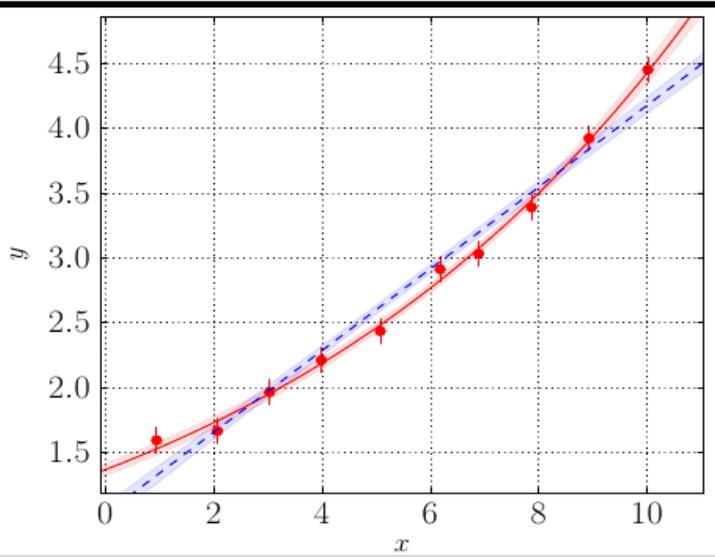


# Auswertung von Messdaten im Praktikum

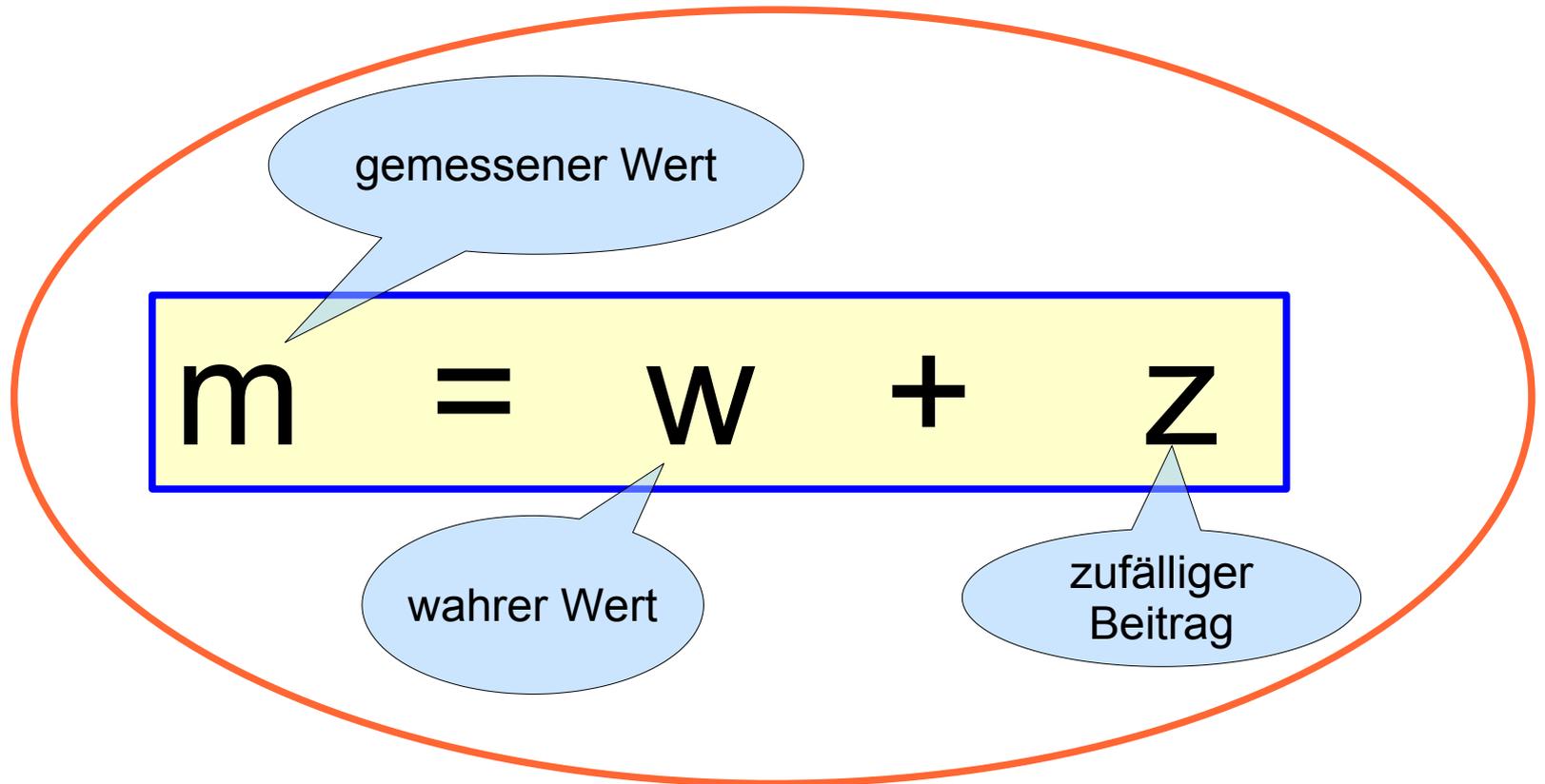
Prof. Dr. Günter Quast, Dr. Hans-Jürgen Simonis

WS 20/21

Fakultät für Physik  
Institut für Experimentelle Kernphysik



# Modell einer Messung



## **Z kann viele Ursachen haben:**

- zufälliger Beitrag zum Messwert („Rauschen“) → „statistische Unsicherheit“
- Genauigkeit des verwendeten Messinstruments → „systematische Unsicherheit“
- mitunter gibt es auch eine Unsicherheit auf den „wahren“ Wert, den man oft z zuschlägt → „theoretische Unsicherheit“
- Fehler im Messprozess – **sollten nicht passieren !**

$$\Rightarrow Z = Z_{\text{stat}} + Z_{\text{sys}} + Z_{\text{theo}}$$

# Darstellung von Messergebnissen

Messung einer physikalischen Größe  $x$  bedeutet,  
experimentell die **Maßzahl** zur Maßeinheit zu ermitteln:

$$G = (G) \cdot [G]$$

Dazu gehört auch die Angabe der **Messunsicherheit  $\Delta G$**

$$G = (x \pm \Delta x) \cdot \langle \text{Maßeinheit} \rangle$$

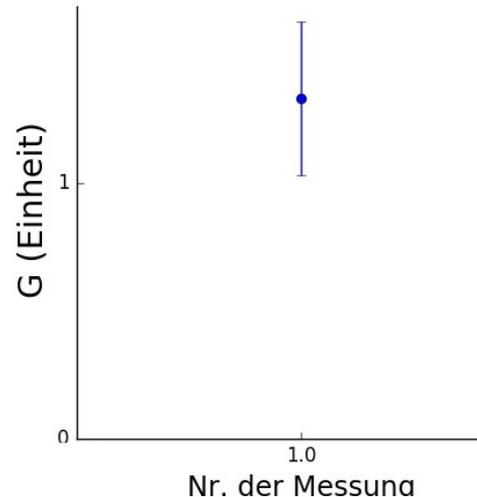
Üblich ist auch die Angabe einer **relativen Meßunsicherheit**:

$$G = (x) \cdot \langle \text{Maßeinheit} \rangle \pm \Delta x / x$$

$\Delta x / x$  meist in % angegeben,  
aber ggf. auch auch ‰ oder  
ppM („parts per Million“)

## Grafische Darstellung:

Koordinatensystem mit  
Messpunkt und „Fehlerbalken“



# Beispiel einer Messung

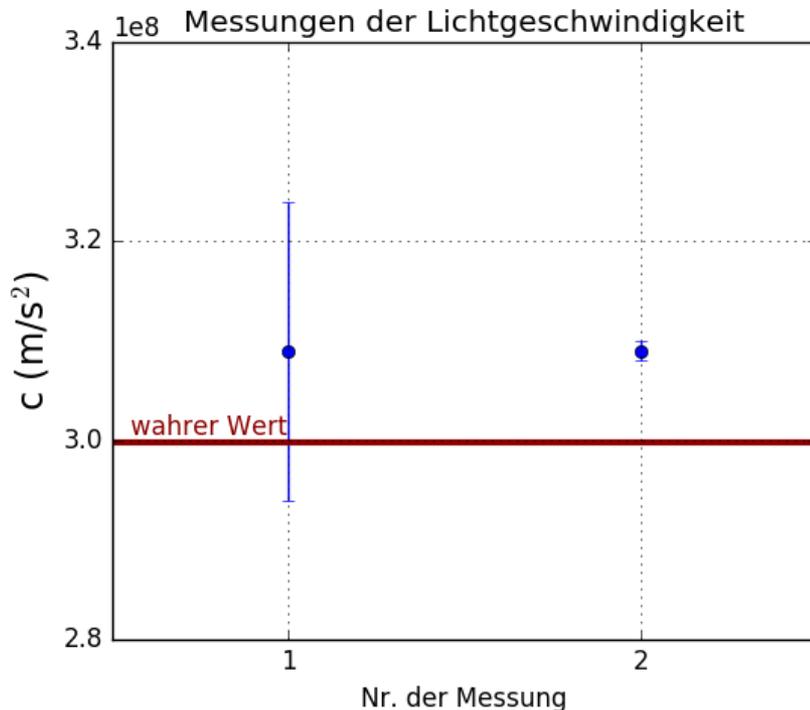
Entscheidend für Bewertung eines Ergebnisses ist die **Messunsicherheit** !

**Beispiel:** Die Lichtgeschwindigkeit ist bereits sehr genau bekannt,  
„Literaturwert“ :  $c = 2.99792458 \cdot 10^8$  m/s

Eine Messung  $c = (3.09 \pm 0.15) \cdot 10^8$  m/s ist in **Übereinstimmung**

Eine Messung  $c = (3.09 \pm 0.01) \cdot 10^8$  m/s wäre dagegen im **Widerspruch**

grafische Darstellung mit erzeugendem Script in der Sprache python



```
import numpy as np, matplotlib.pyplot as plt
# die Messdaten
c_m=[3.09e8, 3.09e8]
c_e=[0.15e8, 0.01e8]
c_w=2.99792458e8
# grafische Darstellung
plt.errorbar([1, 2], c_m, yerr=c_e, fmt='bo')
plt.axhline(c_w, color='darkred', linewidth=3)
plt.text(0.55, 3.005e8, 'wahrer Wert',
color='darkred')
plt.ylabel("c (m/s2)", size='x-large')
plt.xlabel("Nr. der Messung")
plt.title("Messungen der
Lichtgeschwindigkeit")
# (... + einige Verschoenerungen ...)
plt.show()
```

**Angabe einer physikalischen Größe ohne Messunsicherheit ist wertlos!**

# Signifikante Stellen eines Ergebnisses

**Nicht alle Stellen, die ein Rechner ausspuckt, sind relevant !**

**Regel:** Im Praktikum sollten die Messunsicherheiten („Fehler“) auf eine signifikante Stelle gerundet werden.

Die letzte Stelle des Messwerts hat die gleiche Größenordnung

**(Ausnahme:** wenn ein Zwischenergebnis mit Unsicherheit weiter verwendet werden soll, so wird mindestens eine signifikante Stelle mehr mitgenommen)

## Beispiel:

Das numerische Ergebnis einer Messung der Erdbeschleunigung sei

$$g = (9.8234 \pm 0.02385) \text{ m/s}^2$$

→ Angabe  **$g = (9.82 \pm 0.02) \text{ m/s}^2$**   
bzw.  **$g = 9.82 \text{ m/s}^2 \pm 0.2 \%$**

**Hinweis:** übersichtliche Schreibweise ist wichtig !

$m = (0.0000082 \pm 0.0000003) \text{ kg}$  ist zwar korrekt, aber schwer lesbar

$m = 0.0000082 \text{ kg} \pm 0.3 \text{ mg}$  dito

**$m = (8.2 \pm 0.3) \times 10^{-6} \text{ kg}$  oder  $m = (8.2 \pm 0.3) \text{ mg}$  ist viel besser !**

# Fehleranalyse

Zu jeder Messung gehört eine **Fehleranalyse**:  
Ursache und Größe von Messabweichungen

- Ursachen:**
- 1) Fehlerhafte Bedienung von Messgeräten (z.B. falsch kalibriert)
  - 2) Irrtum beim Protokollieren oder der Auswertung (z.B. Zahlendreher)
  - 3) Messverfahren oder Messbedingung ungeeignet

**Grobe Abweichungen** durch sorgfältiges Experimentieren und Kontrolle  
(möglichst durch eine zweite Person) vermeiden !

Grob fehlerhafte Werte einer Messreihe werden nicht weiter verwendet.

*Eine nicht zu unterschätzende Fehlerquelle ist eine mangelnde Objektivität des Experimentators. Oft entstehen falsche Messresultate auch dadurch, dass der Experimentator das Resultat, das er haben will, aus unzureichenden Daten herausliest oder sogar Daten manipuliert.*

kleiner Exkurs: der Fall „Schön“ (DPA-Meldung vom 11.06.2004)

Die **Universität Konstanz entzieht dem Physiker Jan Hendrik Schön seinen Dokortitel!**

Die Universität bezieht sich ausdrücklich nicht auf Fehler in seiner Doktorarbeit, sondern stuft **Datenmanipulationen** während seiner späteren Forschertätigkeit an den Bell-Labs in den USA als **wissenschaftlich unwürdiges Handeln** ein. Das baden-württembergische Universitätsgesetz lässt einen Titelentzug auch auf Grund späteren unwürdigen Verhaltens zu.

# Systematische Unsicherheiten

„**Systematische Fehler**“ zeigen bei identischen Messbedingungen immer um den gleichen Betrag in die gleiche Richtung.

*Sie können durch Messwiederholung weder erkannt noch eliminiert werden. Sie beeinflussen alle unter gleichen Bedingungen erfolgten Messungen in gleicher Weise. Sie sind erfassbar durch Variation der Messmethode oder der Messbedingungen.*

<b>Ursachen</b>	<b>Beispiele</b>
Fehlerhafte Meßgeräte	Eich- oder Justierfehler, Drift,...
Umwelteinflüsse	Temperatur, Druck,...
Rückwirkung der Meßgeräte	Innenwiderstand, Verformung, ..
Unzulänglichkeit des Experimentators	
Gültigkeitsgrenzen der phys. Gesetze	

**Erkannte systematische Probleme können und müssen korrigiert werden!**

**Beispiel:** Temperatureausdehnung eines Maßstabes, geeicht bei 20°, verwendet bei 30°, relative Temperatureausdehnung  $\alpha = 0.0005 / \text{K}$

→ **Korrektur  $L' = L \cdot [1 + 0.0005 \cdot (30-20)] = 1.005 \cdot L$**

Aus mess- oder rechentechnischen Gründen nicht erfassbare systematische Unsicherheiten müssen **abgeschätzt** werden

# Statistische (zufällige) Messunsicherheiten

„**Statistische Fehler**“ beeinflussen Messergebnisse trotz identischer Bedingungen unterschiedlich in Betrag und Vorzeichen.

*Sie sind zufällig in dem Sinne, dass ihre Ursachen im Einzelnen nicht verfolgt werden können - „Zufall“ durch Unkenntnis in der klassischen Physik bzw. als Eigenschaft des Messprozesses in der Quantenphysik.*

*Statistische Fehler sind unvermeidbar und werden mit mathematischen Methoden der Stochastik und Statistik behandelt.*

Subjektive Ursachen	Objektive Ursachen
Parallaxenfehler	Äußere Einflüsse (p, T)
Skaleninterpolation	Statistische Messgröße (Rauschen)
Reaktionsvermögen	

**Zufällige Messunsicherheiten** lassen sich durch **Messwiederholung** und **Mittelung** reduzieren.

Bei  $n$  Wiederholungen einer Messung mit Einzelunsicherheit  $\Delta x$

gilt für die Unsicherheit des Mittelwerts  $\Delta \bar{x} = \frac{\Delta x}{\sqrt{n}}$

# Wie bestimmt man die Werte der Messunsicherheiten ? Und wie sind sie zu interpretieren ?

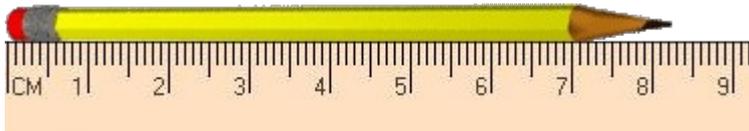
- einmalige Messungen  
→ Fehler abschätzen
- wiederholte gleichartige Messungen  
→ statistische Auswertung
- aus Messgrößen berechnete Ergebnisse  
→ Fehlerfortpflanzung

# Beispiele: einmalig gemessene Größen

Bei einmalig gemessenen Größen schätzt man die Unsicherheit.

Grobe Richtlinien:

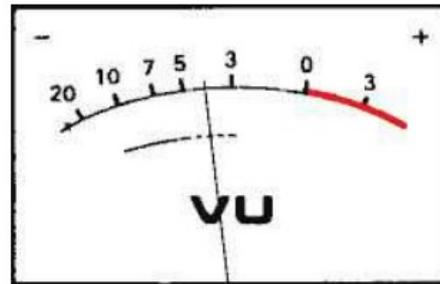
a) bei fein unterteilter Skala



→  $\pm 0.5 \cdot (\text{Intervallbreite})$

$$\underline{L = (82 \pm 0.5) \text{ mm}}$$

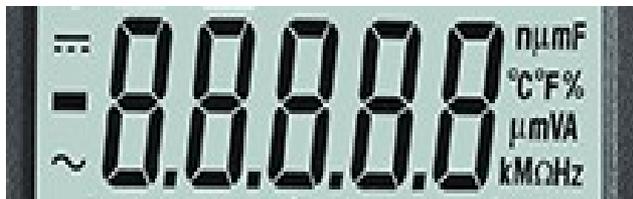
b) bei grob unterteilter Skala



→  $\pm 0.1 \cdot (\text{Intervallbreite})$

$$\underline{U = (4.0 \pm 0.2) \text{ V}}$$

c) digitale Skala



→  $\pm 0.5 \cdot \text{letzte Anzeigestelle}$   
⊕ Geräteklasse (lt. Datenblatt)  
⊕ evtl. Anzeigefluktuationen

# Beispiel: Messreihen

Bei Messreihen ( mit N Werten  $x_i$  ) führt man eine **statistische Analyse** durch.

In der Sprechweise der Statistik:

Eine Messreihe stellt eine Stichprobe aus der Menge der möglichen Messwerte („Grundgesamtheit“) dar. Die Häufigkeitsverteilung der Einzelmessungen nähert sich mit steigender Stichprobengröße der Verteilungsdichte der Grundgesamtheit an.

Unter recht schwachen Voraussetzungen (s. **zentraler Grenzwertsatz**) ist die zu Grunde liegende Verteilungsdichte in der Praxis häufig die **Normal-** oder **Gauß-Verteilung**:

Bestwert oder **Erwartungswert**, auch Mittelwert :

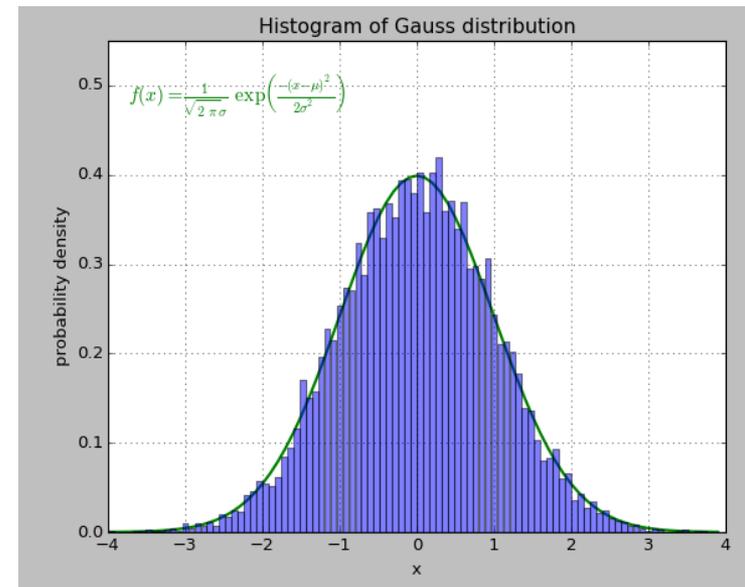
$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$$

Messabweichung oder **Standardabweichung**:

$$\sigma_x = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2}$$

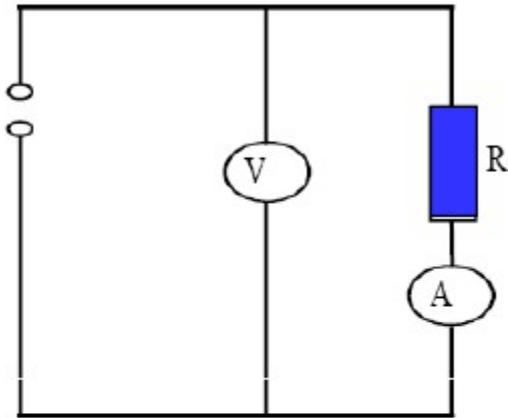
Standardabweichung des Mittelwerts:

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{N}}$$



script [animated\\_Gauss.py](#)

# praktisches Beispiel: Messung eines Widerstands R



$$R = \frac{U}{I}$$

U \ V	I \ mA	R=U/I (Ω)
7,8	35	222.86
15,6	65	240.00
23,4	78	300.00
31,3	126	248.41
39,0	142	274.65
46,9	171	274.27
54,7	194	281.96
62,6	226	276.99
78,3	245	319.59
86,0	258	333.33
87,6	258	339.53
93,6	271	345.39
101,6	277	366.79
109,6	284	385.91
118,0	290	406.90

➔

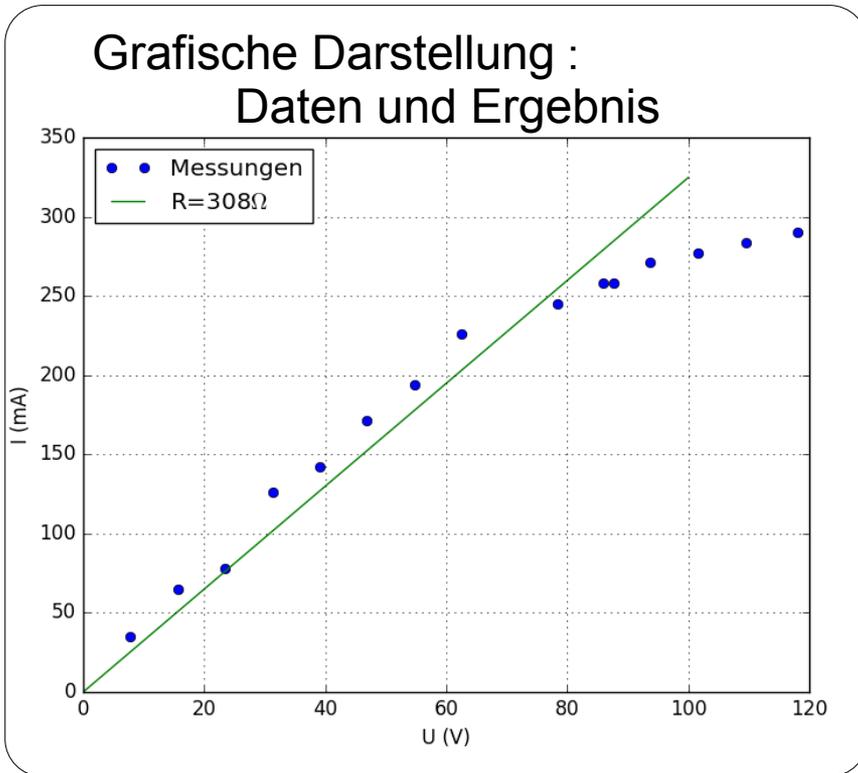
N = 15 Messungen

# praktisches Beispiel: Messung eines Widerstands R (2)

Nach Kochrezept ausrechnen:  $R = \frac{1}{N} \sum R_i = 307.77 \Omega$ ,  $\Delta R = \sqrt{\frac{\sum (R_i - R)^2}{N(N-1)}} = 14 \Omega$

**Aber bei genauerem Hinsehen:**

$$\Rightarrow R = (308 \pm 14) \Omega$$



**Hier ist etwas faul ... ?!**

**Mittelung nicht verträglicher Werte ist unsinnig !** (Hier: Erwärmung des Widerstands führt zu Abweichungen vom Ohm'schen Gesetz)

*Script zur Erzeugung der Grafik*

```
import numpy as np, matplotlib.pyplot as plt

U=[7.8,15.6,23.4,31.3,39.0,46.9,54.7,62.6,78.3,86.0,
  87.6,93.6,101.6,109.6,118.0]
I=[35,65,78,126,142,171,194,226,245,258,258,271,
  277,284,290]

plt.plot(U, I, 'bo', label = "Messungen")
plt.xlabel("U (V)")
plt.ylabel("I (mA)")
x=np.arange(0, 140., 100)
plt.plot(x, 1/0.308*x,'g-',label = "R=308\Omega")
plt.legend(loc='best')
plt.grid()
plt.show()
```

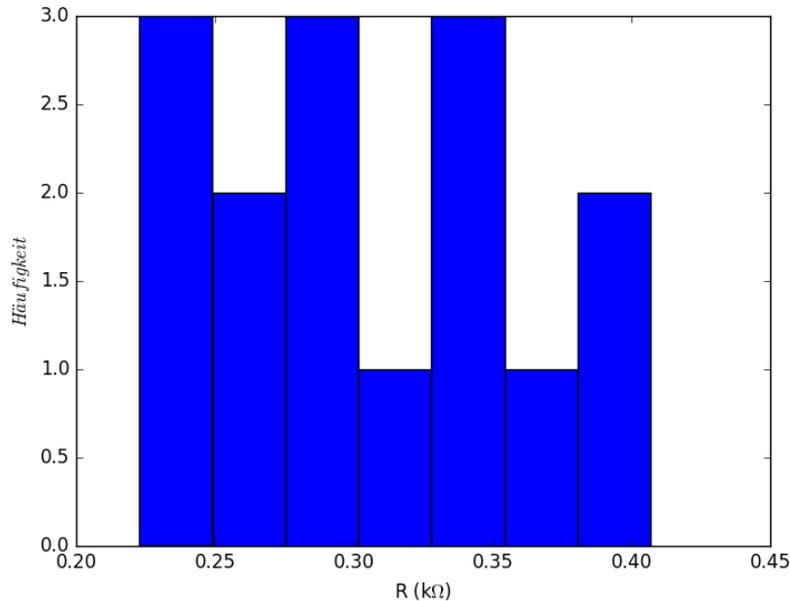
# Häufigkeiten und Histogramm-Darstellung

Auswertung umfangreicher Messreihen:

→ Einteilung der  $N$  Messungen in  $k$  „Klassen“ (z.B. Intervalle 1, ...,  $k$ ) und Häufigkeit des Vorkommens auftragen.

**Histogrammdarstellung:** „Balkendiagramm“ der Häufigkeiten  $H_i, i=1, \dots, k$

Beispiel: Widerstandsmessungen von eben als Häufigkeitsdiagramm (Histogramm)



$$\text{Häufigkeiten: } \sum_{i=1}^k H_i = N$$

$$\text{relative Häufigkeiten: } h_i = \frac{H_i}{N} \Leftrightarrow \sum h_i = 1$$

$$\text{Mittelwert der Messgrößen: } \bar{x} = \sum h_i x_i$$

$$\text{Messunsicherheit: } \sigma_x = \sqrt{\sum h_i (x_i - \bar{x})^2}$$

```
import numpy as np, matplotlib.pyplot as plt
```

```
U=np.array([7.8,15.6,23.4,31.3,39.0,46.9,54.7,62.6,  
            78.3,86.0,87.6,93.6,101.6,109.6,118.0])
```

```
I=np.array([35,65,78,126,142,171,194,226,245,258,  
            258,271,277,284,290])
```

```
R=U/I
```

```
plt.hist(R,7)
```

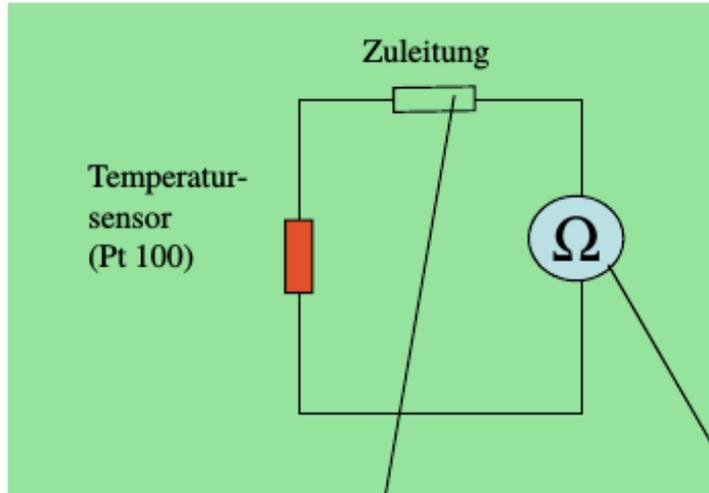
```
plt.xlabel("R ($\Omega$)")
```

```
plt.ylabel(r"$H$ aufigkeit$")
```

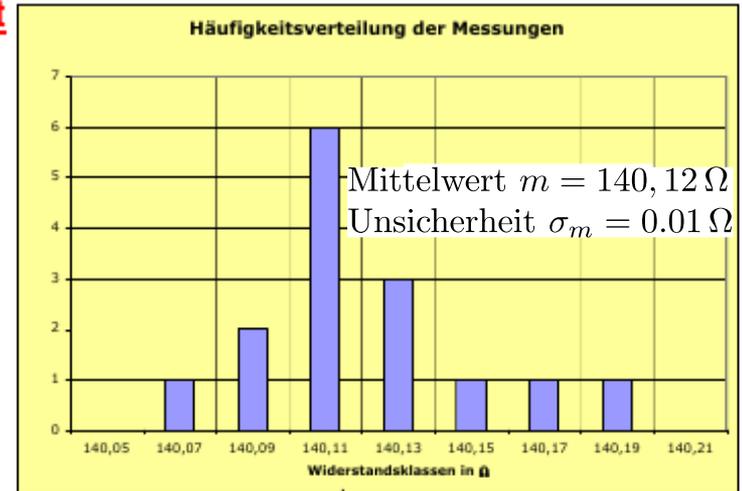
```
plt.show()
```

( s. auch Script  
Histogram.py  
aus CgDA)

# Anwendungsbeispiel: wiederholte Widerstandsmessung



Klasse/ $\Omega$	Häufigkeit
140.05	0
140.07	1
140.09	2
140.11	6
140.13	3
140.15	1
140.17	1
140.19	1
140.21	0
<b>Summe</b>	<b>= 15</b>



**Systematische Abweichung:  
Zuleitungswiderstand  $1,00 \Omega$**

**Herstellergarantie  $0,05 \Omega$**

**statistische Unsicherheit  $0,01 \Omega$**

**Messergebnis:  $140,12 \Omega$**

**Korrektur: Messergebnis - Abweichung  
 $139,12 \Omega$**

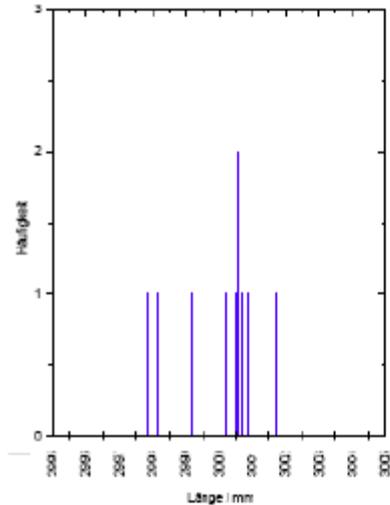
**Endergebnis:  $(139,12 \pm 0,01 \pm 0,05) \Omega$**

**Falls die Gesamtunsicherheit gefordert wird: quadratische Addition**

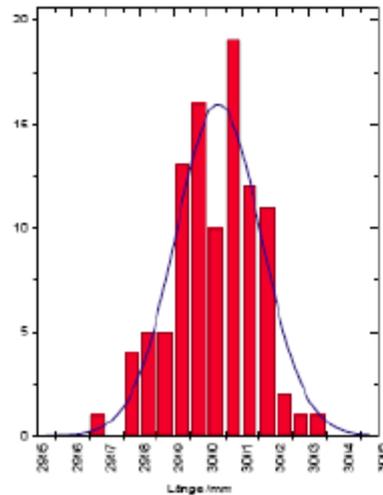
# Grenzverteilung

Mit zunehmender Anzahl von Messungen nähert sich die **Häufigkeitsverteilung** einer **kontinuierlichen Grenzverteilung** an.

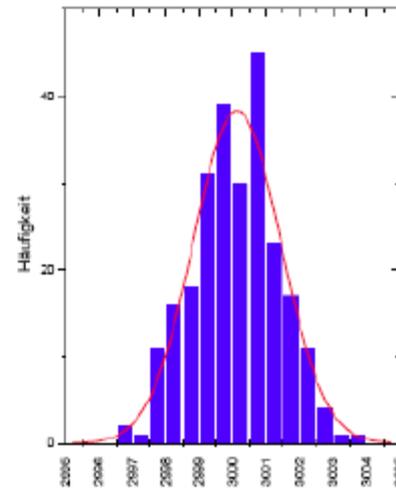
Histogramm (10 Messungen)



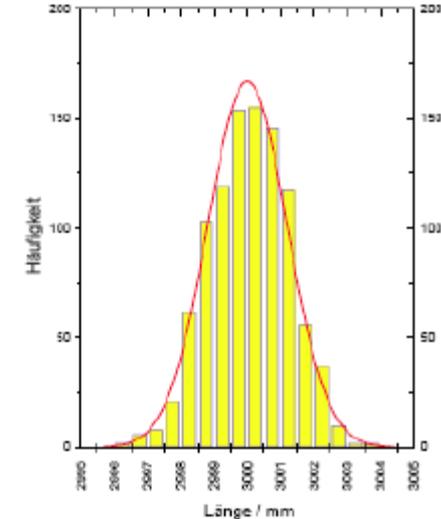
100 Messungen



Histogramm (250 Messungen)



Histogramm (1000 Messungen)

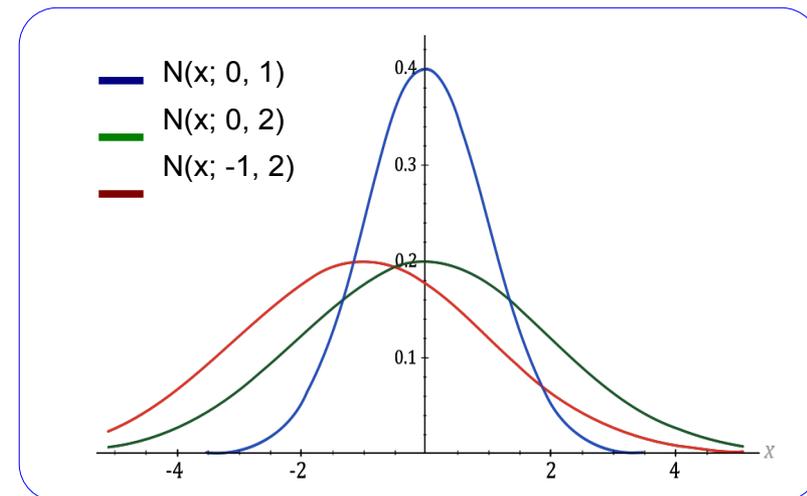


Unter recht schwachen Annahmen (Lyapunov-Bedingung) ist die **Grenzverteilung** der statistischen Abweichungen

die **Gauß-Verteilung**:

$$N(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

(zentraler Grenzwertsatz der Statistik)



# Eigenschaften der Gauß- oder Normal-Verteilung

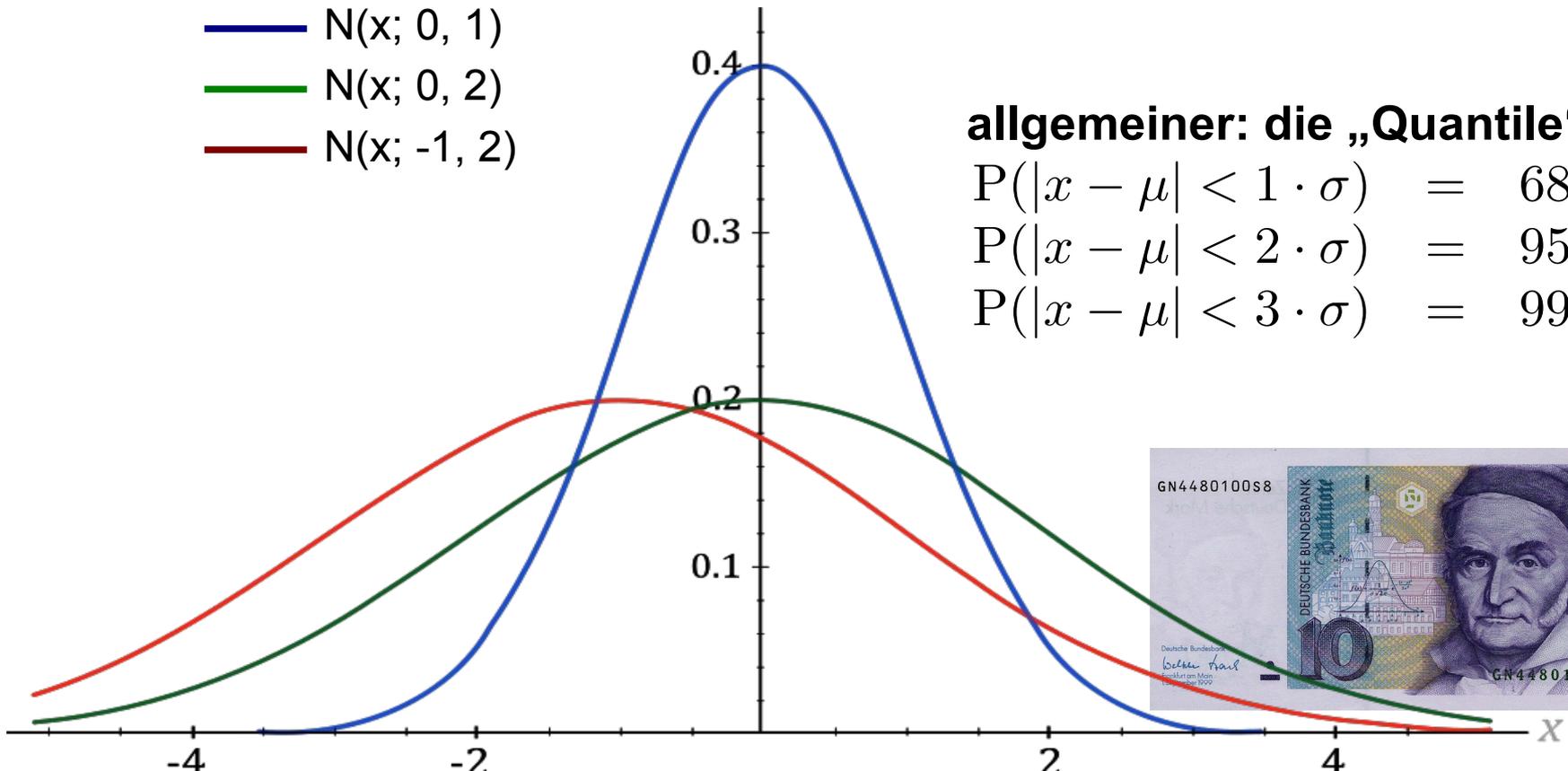
$$N(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

**Mittelwertwert**  $E[x] = \mu$

**Standardabweichung**  $\sigma$   
(Maß für die Breite, entspricht der Messunsicherheit  $\Delta x$  von eben)

68,3% aller Messungen liegen im Intervall  $[\mu - \sigma, \mu + \sigma]$

- $N(x; 0, 1)$
- $N(x; 0, 2)$
- $N(x; -1, 2)$



**allgemeiner: die „Quantile“**

$$\begin{aligned} P(|x - \mu| < 1 \cdot \sigma) &= 68.26\% \\ P(|x - \mu| < 2 \cdot \sigma) &= 95.45\% \\ P(|x - \mu| < 3 \cdot \sigma) &= 99.73\% \end{aligned}$$

# Gewichteter Mittelwert

Wenn eine Größe mehrfach (Werte  $x_i$ ) mit unterschiedlichen Unsicherheiten ( $\sigma_i$ ) bestimmt wurde, so bildet man einen „**gewichteten Mittelwert**“.

Die **Gewichte** sind dabei die **quadratierten Kehrwerte der Unsicherheiten**:

$$\bar{x} = \frac{\sum w_i x_i}{\sum w_i} \text{ mit } w_i = 1/\sigma_i^2$$

**Präzisere Messungen erhalten das größere Gewicht.**

*Überlegen Sie, was die quadratische Wichtung für Ihre Arbeit bedeutet, wenn Sie eine Größe mit dem dreifachen der bisher bekannten Unsicherheit bestimmen.*

Die **Unsicherheit** des gewichteten Mittelwerts ergibt sich folgendermaßen:

$$\sigma_{\bar{x}} = 1 / \sqrt{\sum w_i} \quad \left( \Leftrightarrow \frac{1}{\sigma_{\bar{x}}^2} = \sum \frac{1}{\sigma_i^2} \right)$$

(Beweis z.B. mit Hilfe der Fehlerfortpflanzung,  
s. u., bzw. Vorl. CgDA)

python Script:

```
import numpy as np

w = 1/sx**2
sumw = np.sum(w)
mean = np.sum(w*x)/sumw
smean = np.sqrt(1./sumw)
```

# Fehlerfortpflanzung

Wenn eine Größe  $y$  nicht direkt messbar ist, sondern als Funktion von anderen (Mess-)Größen abhängt, also  $y = f(x_1, \dots, x_n)$ , dann wendet man das **Gauß'sche Fehlerfortpflanzungsgesetz** an:

$$\sigma_y^2 = \sum_{i=1}^n \left( \frac{\partial y}{\partial x_i} \right)^2 \sigma_{x_i}^2$$

Dabei wird die **statistische Unabhängigkeit** der Größen  $x_i$  vorausgesetzt. Für Nicht-lineare Funktionen  $f$  muss die Gültigkeit der **Taylor-Näherung** um die Bestwerte der  $x_i$  innerhalb der Unsicherheiten  $\sigma_{x_i}$  gewährleistet sein. Wenn diese Voraussetzungen nicht gegeben sind, müssen andere Methoden angewandt werden (Berücksichtigung der Kovarianz-Matrix oder Monte-Carlo-Methode, s. Vorl. CgDA).

Spezialfälle:

$$y = x_1 + x_2 \text{ oder } y = x_1 - x_2$$

$$\Rightarrow \sigma_y^2 = \sigma_1^2 + \sigma_2^2$$

Quadrierter *absoluter* Fehler auf die Summe oder Differenz zweier Messungen ist die quadratische Summe ihrer *absoluten* Fehler

$$y = x_1 \cdot x_2 \text{ oder } y = \frac{x_1}{x_2}$$

$$\Rightarrow \left( \frac{\sigma_y}{y} \right)^2 \simeq \left( \frac{\sigma_1}{x_1} \right)^2 + \left( \frac{\sigma_2}{x_2} \right)^2$$

Quadrierter *relativer* Fehler auf das Produkt oder Verhältnis zweier Messungen ist die quadratische Summe ihrer *relativen* Fehler

# Zusammenfassung: Messung

Messwert

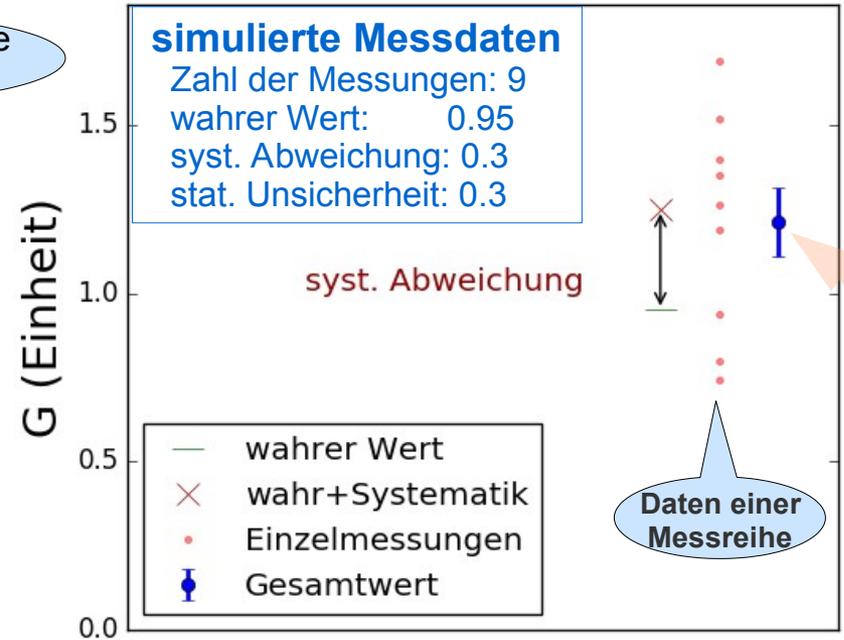
wahrer Wert

zufällige statistische  
Abweichung

$$m = w + Z_{\text{sys}} + Z_{\text{stat}}$$

zufällige systematische  
Abweichung

- systematische Unsicherheit betrifft alle Messwerte in gleicher Weise
- statistische Unsicherheit ist bei jeder Messung anders



Statistische Unsicherheiten können durch Mehrfachmessung und Mittelwertbildung reduziert werden → Zusammengefasst als **Messpunkt mit „Fehlerbalken“**

$$\sigma_{\bar{G}} = \frac{\sigma_G}{\sqrt{N}}$$

Gesamtunsicherheit:  $\sigma_{\text{tot}} = \sqrt{\sigma_{\text{stat}}^2 + \sigma_{\text{sys}}^2} = \sqrt{0.3^2/9 + 0.3^2} = 0.32$

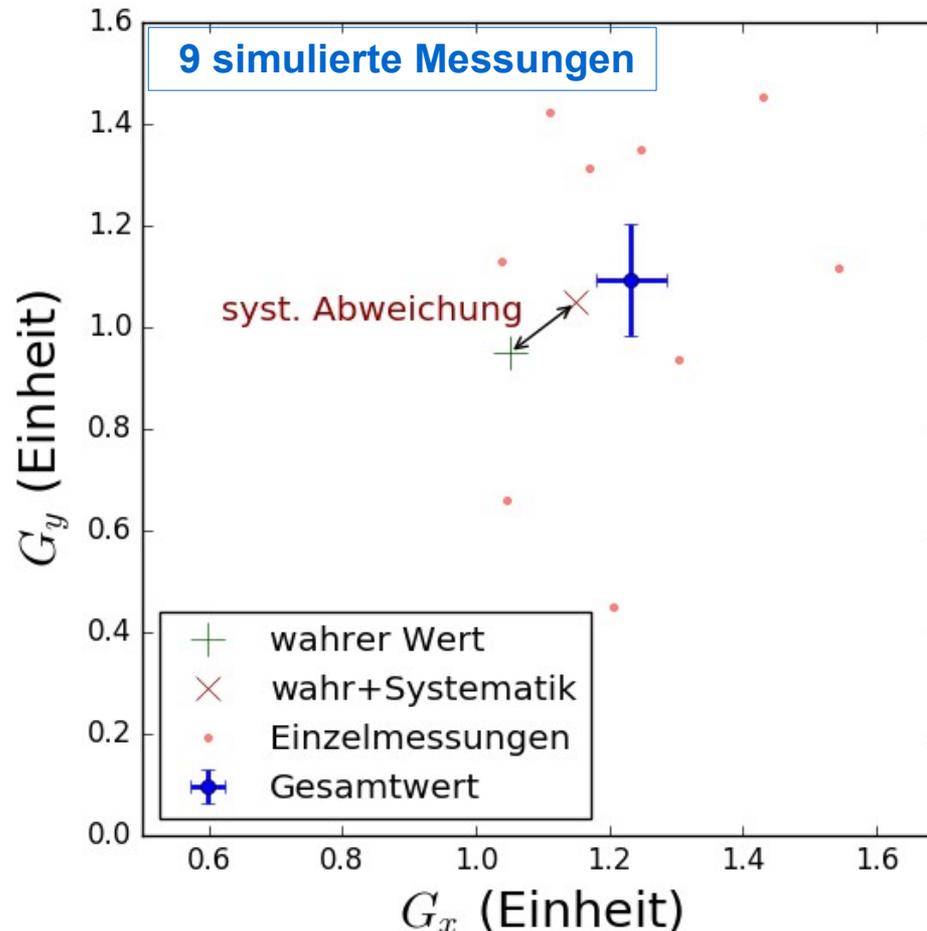
dominiert durch systematische Unsicherheit !

- da verschiedene Unsicherheiten lt. **Fehlerfortpflanzungsgesetz** quadratisch addiert werden, können kleine Beiträge im Endergebnis vernachlässigt werden  
→ **Konzentration auf die dominanten Effekte !**

# Paare von Messungen

Sehr häufig werden Paare von Messungen ( $G_x$ ,  $G_y$ ) aufgenommen.  
Fast immer sind diese Messungen unabhängig –  
dann lassen sich die Überlegungen verallgemeinern:  
**aus dem „Fehlerbalken“ wird ein Fehlerkreuz.**

Mittelwert, Unsicherheit auf den Mittelwert und Fehlerfortpflanzung in jeder Koordinate separat berechnen



Anm.: das Fehlerkreuz zeigt  
nur die statistischen Unsicherheiten

# Datenquellen

## Daten aus Messungen

können in unterschiedlicher Form gewonnen werden:

- Ablesen von analogen und digitalen Messgeräten

Eingabe über Tastatur zur Darstellung / Auswertung

- Datenexport aus digitalen Messgeräten,  
insb. „Datenlogger“ oder Digitaloszilloskope

typischerweise große Datenmengen erfordern Programm-gestütztes Einlesen  
sowie Funktionen zur Darstellung und Signalverarbeitung  
(Maxima / Minima, Periodendauer bzw. Frequenz, Frequenzspektrum, ...)

- Weiterverarbeitung der Ausgabe von Analyseprogrammen

erfordert Funktionen zur Datenübergabe an Programme zur  
finalen Auswertung und Darstellung der Ergebnisse

# (übliche) Datenformate

Messgeräte (auch „Datenlogger“) und einige Handy-Apps (z.B. [phyphox](#)) nutzen einfache Datenformate in Text-Form:

Beispiel-Code

Beispiel PicoScope, „Comma Separated Values“ (CSV)	
Time, Channel A (ms), (V)	} Kopfzeilen mit sog. „ <b>Meta-Daten</b> “
-0.349, -0.000458 -0.348, -0.000458 -0.347, -0.000458 ...	

Üblich sind auch „Tabulator-getrennte“ Dezimalzahlen und - bisweilen – auch Dezimalzahlen mit „**,**“ statt „**.**“ (dann müssen für die Verwendung in python-Programmen Dezimalkommata durch Dezimalpunkte ersetzt werden!)

Durchgesetzt haben sich „beschreibende“ Datenformate, z.B.  
**xml** = „**e**xtensible **m**arkup **l**anguage“ oder  
**json** = „**j**ava **s**cript **o**bject **n**otation“ (≙ **python dictionary**);  
gut durch **python-Module** unterstützt !

Rein „binäre“ Datenformate (also sehr kompakte Darstellungen in maschinenabhängigem digitalem Format) werden wegen ihrer Plattformabhängigkeit heute kaum noch verwendet.

```
# Daten im csv-Format lesen

# Datei zum Lesen öffnen
f = open('AudioData.csv', 'r')

# Kopfzeile(n) lesen
header=f.readline()
print "Kopfzeile:", header

# Daten in 2D-numpy-array einlesen
data = np.loadtxt(f,
                  delimiter=',', unpack=True)
print "-> Anzahl Spalten",
      data.shape[0]
print "-> Datenzeilen",
      data.shape[1]

# Daten in 1D-arrays speichern
t = data[0]
a = data[1]
l = len(a)
```

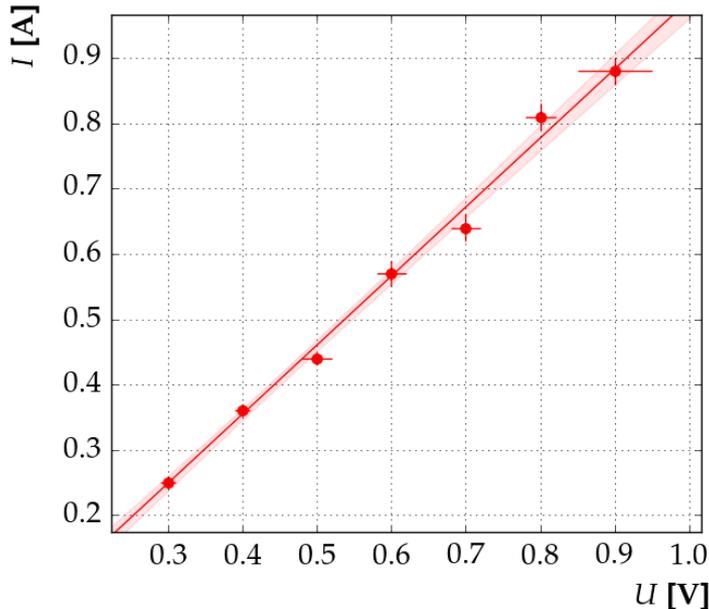
# Modellanpassung oder „Regression“

In der Regel gibt es einen funktionalen Zusammenhang zwischen (Mess-)Größen:

$$y = f(x; p_1, \dots, p_n).$$

Die gesuchten physikalischen Größen stecken dann in den Parametern  $p_1, \dots, p_n$ .

Beispiel von eben ( $I, U$ ):  $I = (1/R) \cdot U$



## in aller Kürze:

Mit numerischen oder analytischen Methoden werden die Parameter  $p_k$  so bestimmt, dass ein vorgegebenes „**Abstandsmaß**“ zwischen den Messpunkten  $y_i$  und den Funktionswerten  $f_i = f(x_i; p_1, \dots, p_n)$  **minimal** wird.

Script: Anpassen von Funktionen an Messdaten

15. November 2013

Funktionsanpassung mit der  $\chi^2$ -Methode

<http://www.ekp.kit.edu/~quast>

## Vorschlag von Gauß:

Summe der kleinsten Fehlerquadrate,

$$S = \chi^2 = \sum_{i=1}^N \frac{(y_i - f(x_i, \{p\}))^2}{\sigma_i^2}$$

wird bzgl. der Parameter  $\{p\}$  minimiert.

# Modellanpassung: Minimieren von $S$

Minimierung von  $S = \chi^2 = \sum_{i=1}^N \frac{(y_i - f(x_i, \{p\}))^2}{\sigma_i^2}$  N Messungen  
k Parameter

- analytisch:  $\frac{\partial S}{\partial p_j} = 0, j = 1, \dots, k$  als notwendige Bedingung für ein Minimum

lösbar in Spezialfällen, z. B. für lineare Probleme  $f(x, \{p\}) = \sum_{j=0}^k p_j f_j(x)$

- i. a. numerisch  
„numerische Optimierung: Algorithmen zur Suche nach dem (einem?)  
Minimum einer Skalaren Funktion im  $k$ -dimensionalen Parameterraum

*In der Praxis werden heute auch für lineare Probleme*

*numerischen Minimierungsmethoden verwendet.*

*(außer in Spezialfällen, z. B. bei zeitkritischen oder  
immer wieder vorkommenden Problemstellungen)*

# Modellanpassung: Beispiel mit einem Parameter

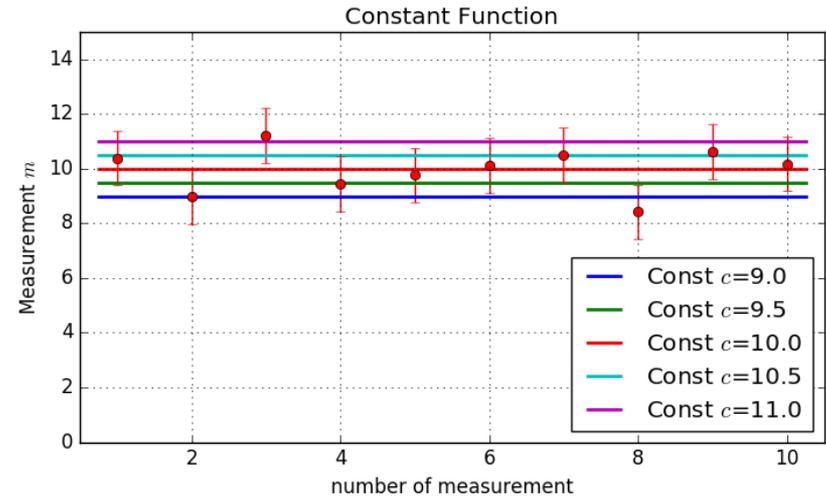
**Mittelwert** von 10 Messungen  $y_i$  mit Unsicherheiten  $\sigma$  entspricht der Anpassung einer konstanten Funktion  $f(x;c)=c$

$$S(c) = \sum_{i=1}^{10} \frac{(y_i - c)^2}{\sigma^2}$$

analytisch:  $0 = \frac{dS}{dc} = \sum_{i=1}^{N=10} \frac{-2(y_i - c)}{\sigma^2}$

$$\Rightarrow \hat{c} = \frac{1}{N} \sum_{i=1}^{N=10} y_i$$

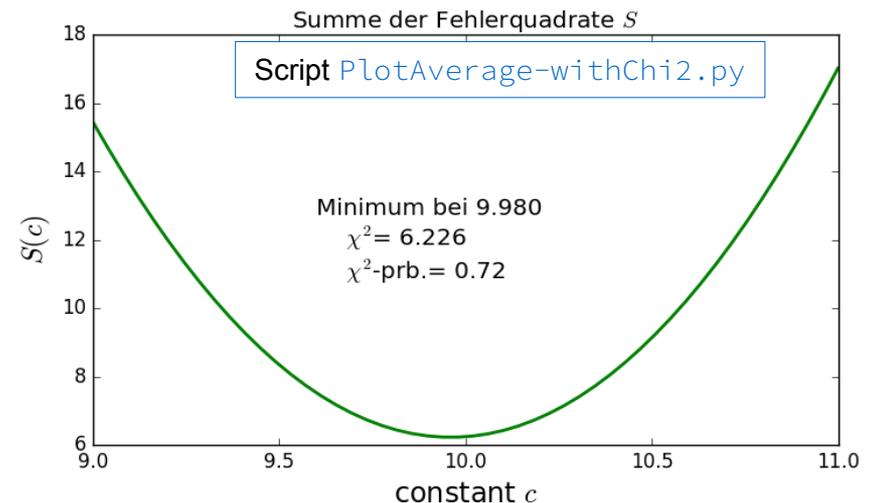
identisch zum „Mittelwert“



„numerisch“:

$$S(c) = \sum_{i=1}^{10} \frac{(y_i - c)^2}{\sigma^2}$$

berechnen und grafisch darstellen



# Spezialfall: Lineare Regression (Geradenanpassung)

$$f(x; p_1, p_2) = p_1 + p_2 x \quad \Rightarrow \quad S = \sum_{i=1}^N \frac{(y_i - p_1 - p_2 x_i)^2}{\sigma_i^2}$$

Nullsetzen der 1. Ableitungen ergibt das Gleichungssystem

$$(1) \quad 0 \stackrel{!}{=} \frac{\partial S}{\partial p_1} = -2 \sum_{i=1}^N \frac{(y_i - p_1 - p_2 x_i)}{\sigma_i^2}$$
$$(2) \quad 0 \stackrel{!}{=} \frac{\partial S}{\partial p_2} = -2 \sum_{i=1}^N \frac{x_i (y_i - p_1 - p_2 x_i)}{\sigma_i^2}$$

mit den Abkürzungen

$$S_1 = \sum_{i=1}^N \frac{1}{\sigma_i^2}, \quad S_x = \sum_{i=1}^N \frac{x_i}{\sigma_i^2} = \bar{x} S_1, \quad S_y = \sum_{i=1}^N \frac{y_i}{\sigma_i^2} = \bar{y} S_1$$
$$S_{xx} = \sum_{i=1}^N \frac{x_i^2}{\sigma_i^2} = \overline{x^2} S_1, \quad S_{xy} = \sum_{i=1}^N \frac{x_i y_i}{\sigma_i^2} = \overline{xy} S_1, \quad D = S_1 S_{xx} - S_x^2$$

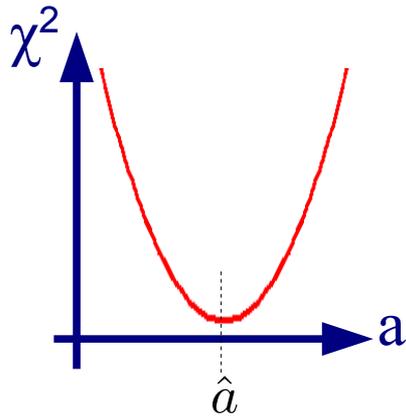
ergibt sich die Lösung:

$$\hat{p}_1 = \frac{S_{xx} S_y - S_x S_{xy}}{D}, \quad \sigma_{p_1}^2 = \frac{S_{xx}}{D},$$
$$\hat{p}_2 = \frac{S_1 S_{xy} - S_x S_y}{D}, \quad \sigma_{p_2}^2 = \frac{S_1}{D}, \quad V_{12} = \frac{-S_x}{D}$$

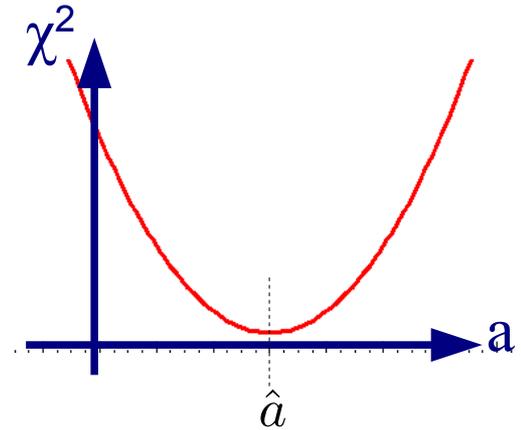
Implementierung in python s. Script `linRegression`

*Diese Formeln waren für Generationen von Studierenden die Basis einer jeden Anpassung („Regression“)*  
*aber: schon die Behandlung von Unsicherheiten in Ordinate **und** Abszisse erfordert numerische Methoden*

# Modellanpassung: Bestimmung der Parameterfehler



Je schärfer das Minimum von  $\chi^2(\mathbf{p})$ ,  
desto kleiner die Parameterfehler:



scharfes Minimum:  
große Krümmung

flaches Minimum:  
kleine Krümmung

→ Parameterunsicherheiten sind umgekehrt proportional  
zu Krümmung(en) von  $\chi^2(\mathbf{p})$  am Minimum

$$\frac{1}{\sigma_{\hat{p}}^2} = \frac{1}{2} \left. \frac{\partial^2 \chi^2(\mathbf{p})}{\partial p^2} \right|_{\hat{p}} \quad \text{bzw.} \quad (V_{\hat{a}}^{-1})_{ij} = \frac{1}{2} \left. \frac{\partial^2 \chi^2(\{\mathbf{p}\})}{\partial p_i \partial p_j} \right|_{\hat{p}_i \hat{p}_j}$$

bei mehreren Parametern .

V: Kovarianzmatrix der Parameterunsicherheiten, s. Vorl. CgDA

# Modellanpassung: das Anpassungspaket kafe

zu **kafe** gibt es eine Anzahl von gut dokumentierten Beispielen, s.

<http://ekpwww.etp.kit.edu/~quast/kafe2/html/doc>

oder

<https://github.com/dsavoiiu/kafe2>

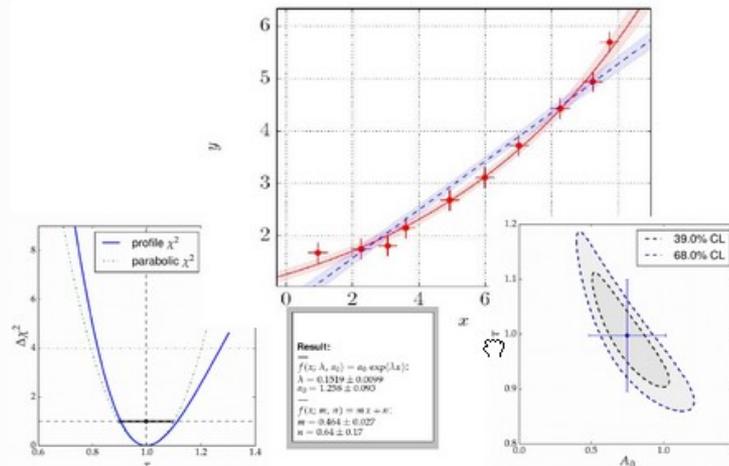
## Welcome to KaFE (Karlsruhe Fit Environment)

**kafe** is a data fitting framework designed for use in undergraduate physics lab courses. It provides a basic *Python* toolkit for fitting models to data as well as visualisation of the data and the model function. It relies on *Python* packages such as `numpy` and `matplotlib`, and uses the *Python* interface to the minimizer *Minuit* contained in the data analysis framework *ROOT*.

### `kafe` Overview

The `kafe` package provides a rather general approach to fitting of a model function to two-dimensional data points with correlated uncertainties in both dimensions. The *Python* API guarantees full flexibility for data input. Helper functions for file-based input and some examples are available for own applications.

Applications range from performing a simple average of measurements to complex situations with both correlated (systematic) and uncorrelated (statistical) uncertainties on the measurements of the x and y values described by a non-linear model function depending on a large



**Graphical output generated with kafe.**

### Anmerkung:

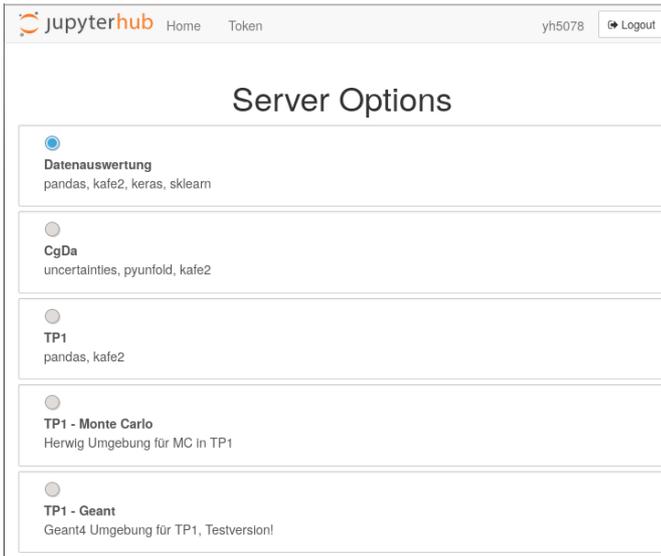
die Berücksichtigung von Unsicherheiten in Ordinaten- und Abszissenrichtung oder die Anpassung von Modellen, die nicht linear von den Parametern abhängen, sind analytisch nicht möglich.

Die **Empfehlung** ist daher, grundsätzlich Anpassungen mit **numerischen Werkzeugen** durchzuführen.



# Arbeiten mit *jupyter*-Notebooks

<https://jupytermachine.etp.kit.edu>



The screenshot shows the JupyterHub interface with the following elements:

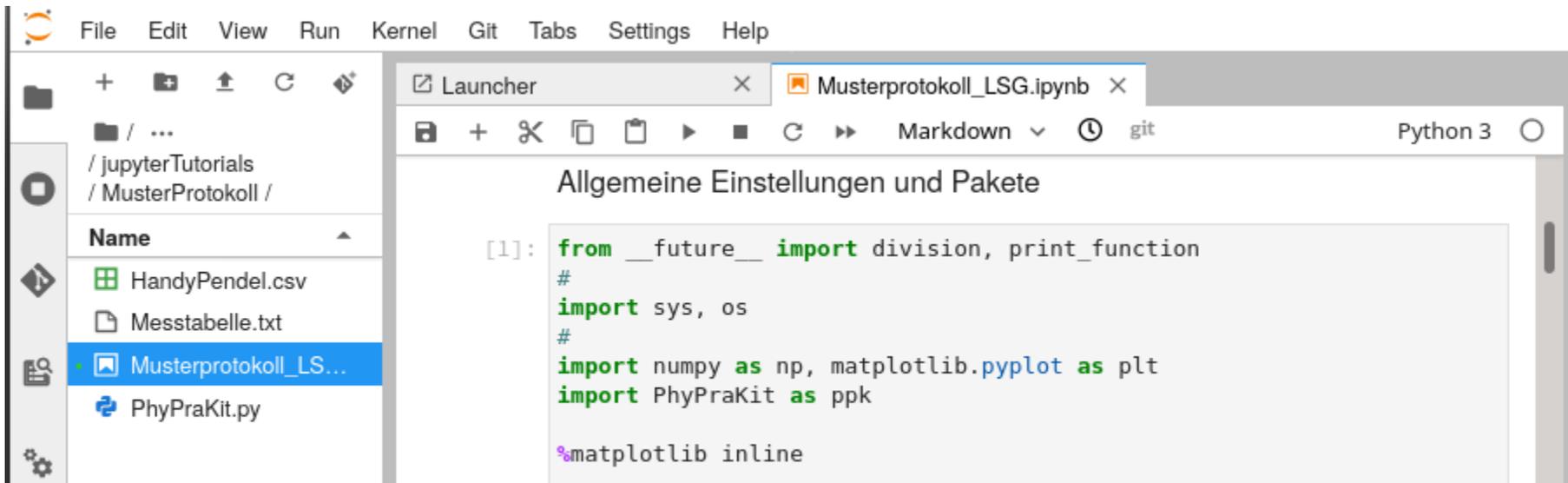
- Header: jupyterhub Home Token yh5078 Logout
- Section: Server Options
- Option 1: **Datenauswertung** (selected) with packages: pandas, kafe2, keras, sklearn
- Option 2: **CgDa** with packages: uncertainties, pyunfold, kafe2
- Option 3: **TP1** with packages: pandas, kafe2
- Option 4: **TP1 - Monte Carlo** with description: Herwig Umgebung für MC in TP1
- Option 5: **TP1 - Geant** with description: Geant4 Umgebung für TP1, Testversion!

Voraussetzung zur **Nutzung des *jupyter*-Servers:**

- KIT-Account und CIP-Pool-Zugang  
s. <https://comp.physik.kit.edu/Account/>

## Beispiele:

- Spickzettel für *jupyter* : [JupyterCheatsheet.ipynb](#)
- Einführung in *Python* : [PythonIntro.ipynb](#)
- Spickzettel für *Python* :  
[PythonCheatsheet.ipynb](#)
- Grundlagen der Statistik: [IntroStatistik.ipynb](#)
- Fehlerrechnung im Physikalischen Praktikum:  
[Fehlerrechnung.ipynb](#)
- Tutorium zu kafe2: [kafe2Tutorial.ipynb](#)



The screenshot shows a Jupyter Notebook interface with the following elements:

- Menu: File Edit View Run Kernel Git Tabs Settings Help
- File browser: / jupyterTutorials / MusterProtokoll /
- File list: HandyPendel.csv, Messtabelle.txt, **Musterprotokoll\_LS...**, PhyPraKit.py
- Launcher window: Musterprotokoll\_LSG.ipynb (Python 3)
- Code cell content:

```
[1]: from __future__ import division, print_function
#
import sys, os
#
import numpy as np, matplotlib.pyplot as plt
import PhyPraKit as ppk

%matplotlib inline
```

# Anhang

Praktische Beispiele

# Beispiel: Rohdaten einer Wellenform

Time, Channel A  
(ms), (V)

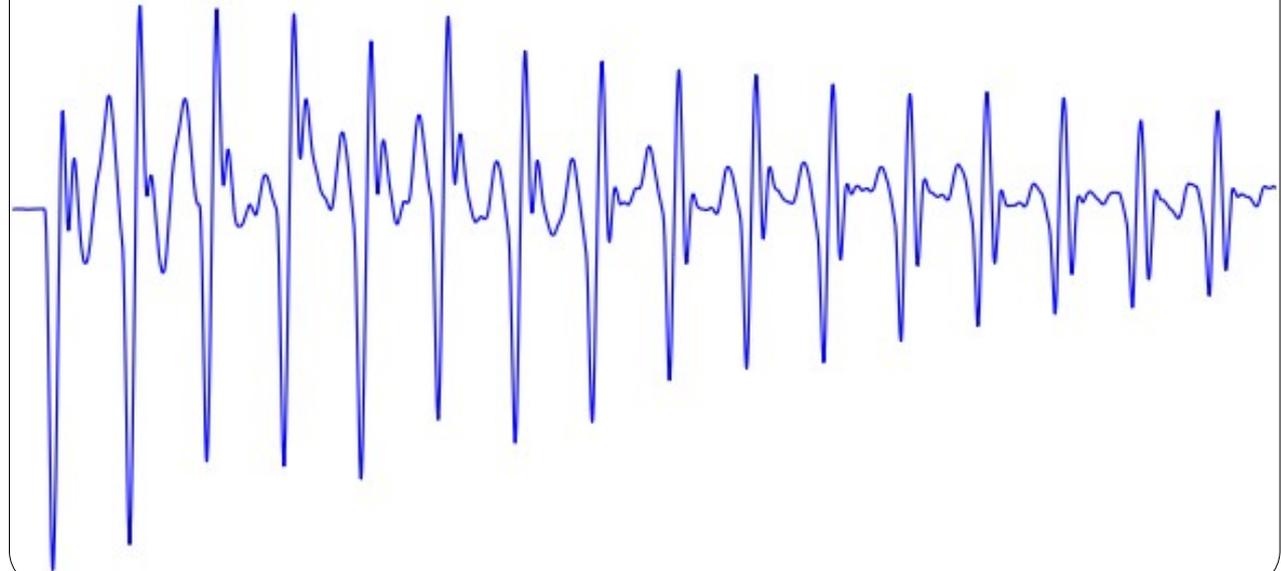
```
-0.34927999, -0.00045778  
-0.34799999, -0.00045778  
-0.34671999, -0.00045778  
-0.34543999, -0.00045778  
-0.34415999, -0.00045778  
-0.34287999, -0.00033570  
-0.34159999, -0.00018311  
-0.34031999, -0.00018311  
-0.33903999, -0.00018311  
-0.33775999, -0.00003052  
-0.33647999, 0.00006104 -  
0.33519999, 0.00006104 -  
0.33391999, 0.00006104 -  
0.33263999, 0.00006104 -  
0.33135999, 0.00021363 -  
0.33007999, 0.00021363
```

. . .

```
9.63983995, 0.02642903  
9.64111995, 0.02655110  
9.64239995, 0.02655110  
9.64367995, 0.02655110  
9.64495995, 0.02655110  
9.64623995, 0.02655110  
9.64751995, 0.02655110  
9.64879995, 0.02642903  
9.65007995, 0.02612384  
9.65135995, 0.02584918  
9.65263995, 0.02557451  
9.65391995, 0.02526933  
9.65519995, 0.02487259
```

7816 Wertepaare exportiert aus  
USB-Oszilloskop PicoScope im  
**csv**-Format (**c**omma **s**eparated **v**alues)

importiert in numpy-arrays `t` und `v`,  
dargestellt mit `plt.plot(t, v)`



Datei  
Wellenform.csv

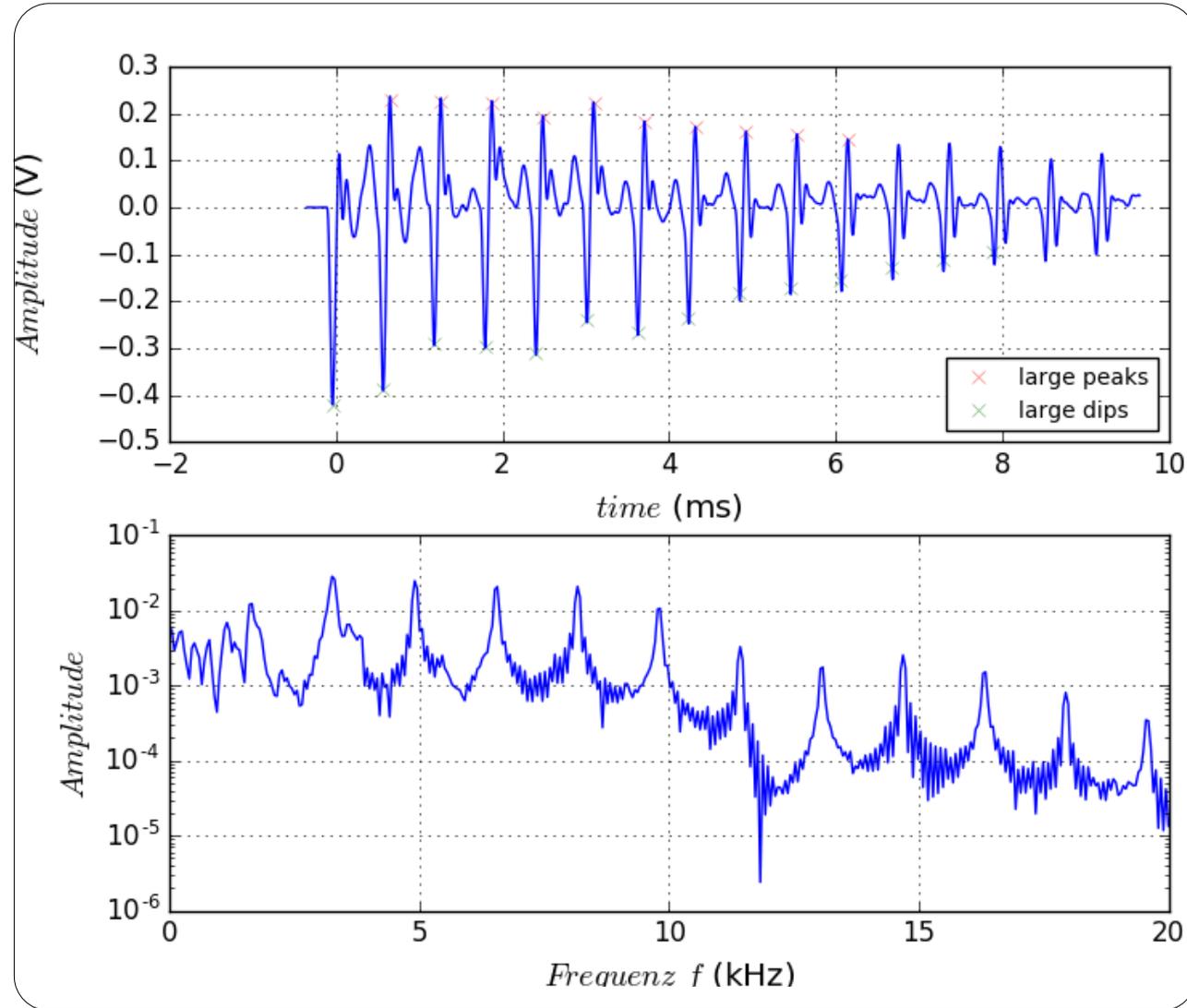
Script:  
test\_readPicoscope.py

# Beispiel Frequenzanalyse

## Amplitudenverlauf

$$a_i = \text{Amplitude}(t_i)$$

(„time domain“)



## Fourier-Transformation:

$$s_k = \frac{1}{n} \sum_i a_i \sin\left(\frac{2\pi}{f_k} t_i\right)$$

$$c_k = \frac{1}{n} \sum_i a_i \cos\left(\frac{2\pi}{f_k} t_i\right)$$

$$A_k = \sqrt{s_k^2 + c_k^2}$$

(„frequency domain“)

Script `test_Fourier.py`

numerisch effizienteres Verfahren: **Fast Fourier Transform, FFT**

# Modellanpassung mit kafe

In der Praxis setzt man Programmpakete ein, die

- Strukturen zur Verwaltung von Daten und deren Fehlern
- Definition von Modellen
- Anpassung mittels numerischer Optimierung
- grafische Darstellung
- Ausgabe der Ergebnisse

bereit stellen.

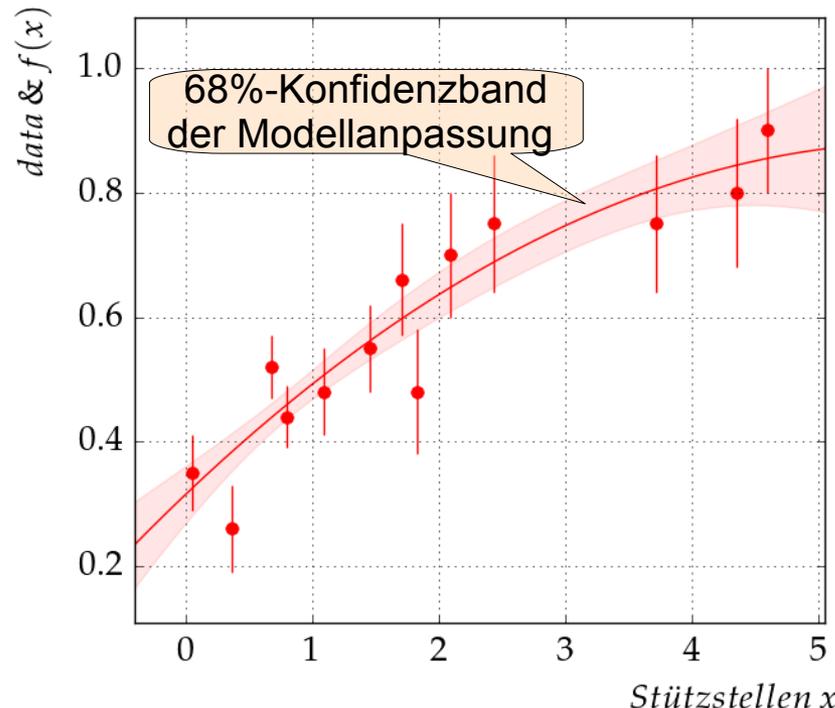
Zuverlässige Konvergenz zum globalen Minimum bei nicht-linearen Problemstellungen erfordert meist das Setzen geeigneter Startwerte!

Beispiel mit dem  
python-Framework

„kafe“:

Anpassung der drei  
Parameter einer qua-  
dratischen Funktion  
an Datenpunkte mit  
Unsicherheiten nur  
in Ordinatenrichtung.

siehe Script  
[fitexample\\_kafe.py](#)



—  $f(x; a, b, c) = a x^2 + b x + c$   
♦ example data

## Fit Info

—  
 $f(x; a, b, c) = a x^2 + b x + c$   
 $a = -0.017 \pm 0.013$   
 $b = 0.193 \pm 0.059$   
 $c = 0.315 \pm 0.046$

# Beispielcode: numerische Anpassung mit kafe

siehe Script [fitexample\\_kafe.py](#)

```
# example fit with kafe
import kafe
from kafe.function_tools import FitFunction, LaTeX, ASCII

# fit function definition (with decorators for nice output)
@ASCII(expression='a * x^2 + b * x + c')
@LaTeX(name='f', parameter_names=('a','b','c'), expression=r'a\,x^2+b\,x+c')
@FitFunction
def poly2(x, a=1.0, b=0.0, c=0.0):
    return a * x**2 + b * x + c

# ----- begin of workflow -----
# set data
xm = [.05,0.36,0.68,0.80,1.09,1.46,1.71,1.83,2.44,2.09,3.72,4.36,4.60]
ym = [0.35,0.26,0.52,0.44,0.48,0.55,0.66,0.48,0.75,0.70,0.75,0.80,0.90]
ye = [0.06,0.07,0.05,0.05,0.07,0.07,0.09,0.1,0.11,0.1,0.11,0.12,0.1]

# create a kafe Dataset
kdata = kafe.Dataset(data=(xm, ym), basename='kData', title='example data')
kdata.add_error_source('y', 'simple', ye) # add uncertainties
kfit=kafe.Fit(kdata, poly2) # create the Fit object from data & fit function
kfit.do_fit() # perform fit
kplot=kafe.Plot(kfit) # create plot object
kplot.axis_labels = [r'$St$'utzstellen \, x $', r'$data\,\&\,f(x)$']
kplot.plot_all() # make plots
kplot.show() # show the plots
```

nur zur  
Verschönerung

der wichtige  
Code für kafe

kafe nutzt das CERN-Paket MINUIT zur numerischen Optimierung

# Modellanpassung: Daten-getriebene Anpassung

Die Methode `kafe.file_tools.build_fit_from_file()` ermöglicht es, einfache Anpassungen **vollständig über eine Datei zu steuern**:

Skript `kafe_fit-from-file.py`

```
import sys, matplotlib.pyplot as plt, kafe
from kafe.file_tools import buildFit_fromFile

# check for / read command line arguments
if len(sys.argv)==2:
    fname=sys.argv[1]
else:
    fname='data.fit'
print '*==* script ' + sys.argv[0]+ ' executing \n',\
      ' processing file ' + fname

# initialize fit object from file and run fit
theFit = buildFit_fromFile(fname)
theFit.do_fit()
thePlot = kafe.Plot(theFit)
thePlot.plot_all( show_info_for=None)

#thePlot.save(fname.split('.')[0]+'pdf')
#theFit.plot_correlations() # eventually contours
thePlot.show() # show everything on screen
```

Mit Hilfe dieses **Universalskripts** müssen

Sie keinen speziellen python-Code schreiben, sondern nur die **Eingabedatei** erstellen !

# example showing fit driven by input file

Datei  
IU-Messungen.fit

```
# - - Meta data for plotting
*BASENAME linearFitExample
*FITLABEL Angepasste Gerade
*TITLE I-U Messungen
*xLabel $U$
*xUnit V
*yLabel $I$
*yUnit A
```

# - - description of data

```
*xData
0.30 0.01
. . .
```

```
0.90 0.05
```

```
*yData
0.25 0.01
. . .
```

```
0.88 0.02
```

# systematic errors: correlated among all measurements

```
*xRelCor 0.005 # relative error of 0.5% on U
```

```
*yRelCor 0.005 # relative error of 0.5% on I
```

# - - fuction to fit

```
*FitFunction
```

```
def fitf(x, a=1., b=0.):
```

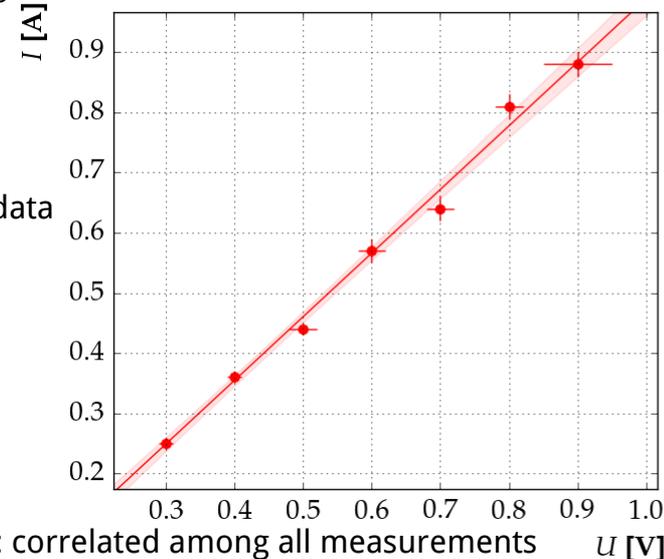
```
~~return a*x +b
```

# initial parameter values and range

```
*InitialParameters
```

```
1. 0.3
```

```
0. 0.1
```



# Kommandozeilen-Werkzeug: **kfitf.py**

es gibt ein Werkzeug für die Kommandozeile, **kfitf.py**  
das in einer .fit -Datei definierte Anpassung ausführt:

kfitf.py

usage: kfit [-h] [-n] [-s] [-c] [--noinfo] [--noband] [-f FORMAT] filename

Perform a fit with the kafe package driven by input file

positional arguments:

filename            name of fit input file

optional arguments:

-h, --help            show this help message and exit

-n, --noplots        suppress output of plots on screen

-s, --saveplot       save plot(s) in file(s)

-c, --contour        plot contours and profiles

--noinfo            suppress fit info on plot

--noband            suppress 1-sigma band around function

-f FORMAT, --format FORMAT

graphics output format, default=pdf

# weitere Beispiele ...

Alle Dateien zu den gezeigten Beispielen (und noch viele mehr)  
finden Sie unter dem Link zur Vorlesung

## Computergestützte Datenauswertung

<http://www.ekp.kit.edu/~quast/CgDA/CgDA.html>

Spezielle **Python-Beispiele zum Start ins Praktikum**

<http://www.ekp.kit.edu/~quast/CgDA/PhysPrakt>

das python-Modul `PhyPraKit.py` enthält

### 1. Einlesen von Daten aus Text-Dateien

`read_data()`      n Spalten mit Meta-Daten (Datum, Autor, ...)  
`read_csv()`      Daten als „Comma Separated Values“  
`read_txt()`      allg. Daten im Text-Format

### 2. Methoden zur Prozessierung von Roh-Daten

Glätten, Suche nach Extrema und Flanken,  
Fouriertransformation, Autokorrelation

### 3. Berechnung des gewichteten Mittelwerts

### 4. Histogramm-Tools

`barstat()`      statistische Information aus Balkendiagramm  
`histstat()`      statistische information aus 1d-Histogramm  
`hist2dstat()`      statistische Information aus 2d-histogramm  
`profile2d()`      "profile plot" für 2d-Daten  
`chi2p_indep2d()`      chi2-Test auf Unabhängigkeit zweier Datensätze

### 5. lineare Regression ( $y = a \cdot x + b$ ), Funktionsanpassung

`linRegression()`      mit der analytischen Formel (nur y-Fehler)  
`kFit()`      numerisch mit kafe, x-, y- und korrelierte Fehler

### 6. Erzeugung von simulierten Datensätzen

### s. Beispiel-Scripts

`test_readtxt.py`

`test_Fourier`  
`test_AutoCorrelation.py`

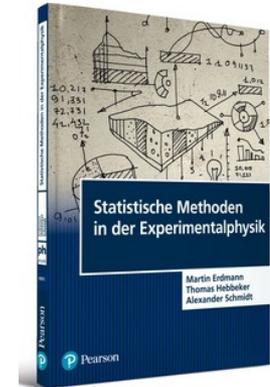
`test_Histogram.py`

`test_linRegression.py`  
`test_kFit.py`

`test_generateData.py`

# Literatur

- M. Erdmann, T. Hebbeker, A. Schmidt, *Statistische Methoden in der Experimentalphysik*, Pearson, 2019
- H.J.Eichler et al., *Das neue Physikalische Grundpraktikum*, Springer Lehrbuch 2016 *digital über die KIT-Bibliothek*
- W.H. Gränicher; *Messung beendet - was nun?*, Teubnerverlag Stuttgart, 1996
- Gerhard Bohm, Günter Zech; *Introduction to Statistics and Data Analysis for Physicists*, Verlag Deutsches Elektronen-Synchrotron 2014 *digital als DESY ebook*
- Eigene Skripte zur Veranstaltung



- Funktionsanpassung
- Software-Paket *kafe*
- *jupyter* Server
- Virtuelle Maschine

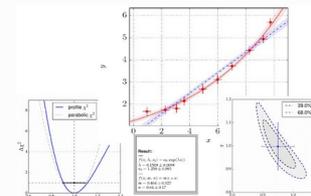
<http://www.ekp.kit.edu/~quast>

## Welcome to KaFE (Karlsruhe Fit Environment)

*kafe* is a data fitting framework designed for use in undergraduate physics lab courses. It provides a basic Python toolkit for fitting models to data as well as visualisation of the data and the model function. It relies on Python packages such as *numpy* and *matplotlib*, and uses the Python interface to the minimizer *Minuit* contained in the data analysis framework *ROOT* or available as a separate python package *iminuit*.

### *kafe* Overview

The *kafe* package provides a rather general



Graphical output generated with *kafe*.

Script: Anpassen von Funktionen an Messdaten

26. Februar 2015

Funktionsanpassung mit der  $\chi^2$ -Methode

Script: Einsatz von Virtuellen Maschinen

29. August 2014

VIRTUELLE MASCHINE ZUR PHYSIK

**Danke für Ihre Aufmerksamkeit**

Wir wünschen Ihnen

Experimentierfreude und ein

spannendes, erkenntnisreiches Praktikum