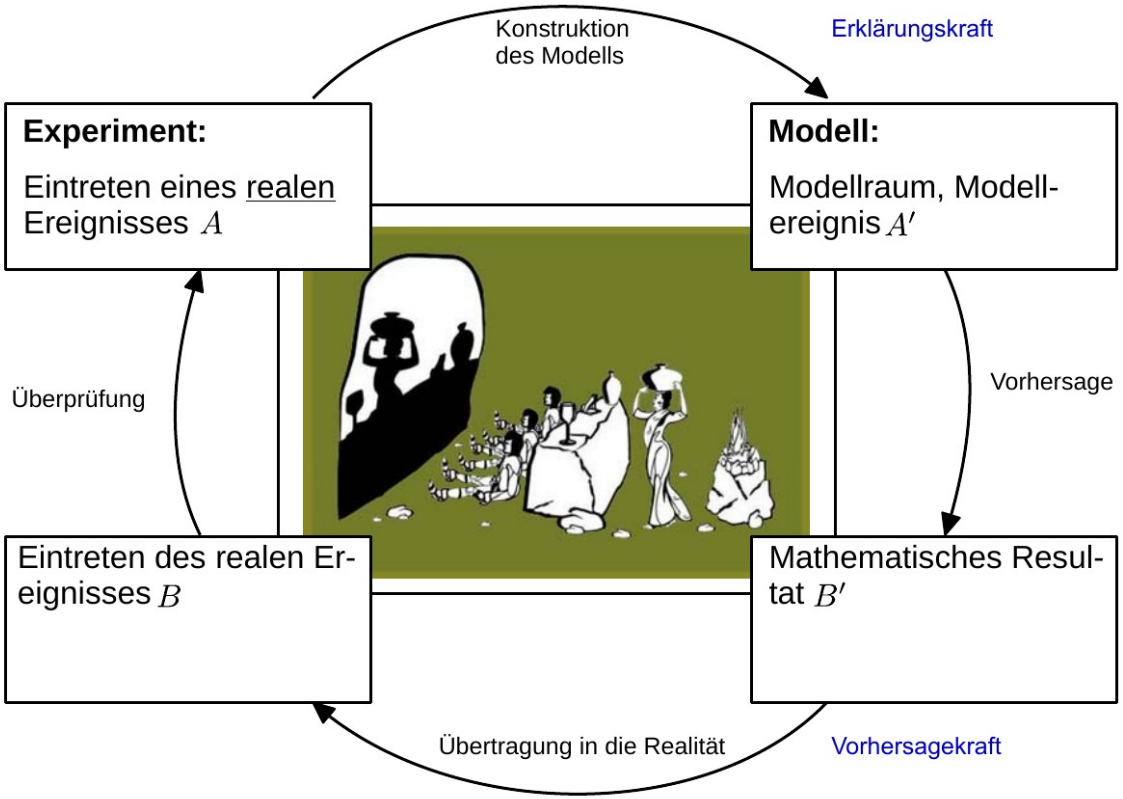
Einführung in die Datenanalyse im Praktikum



Roger Wolf

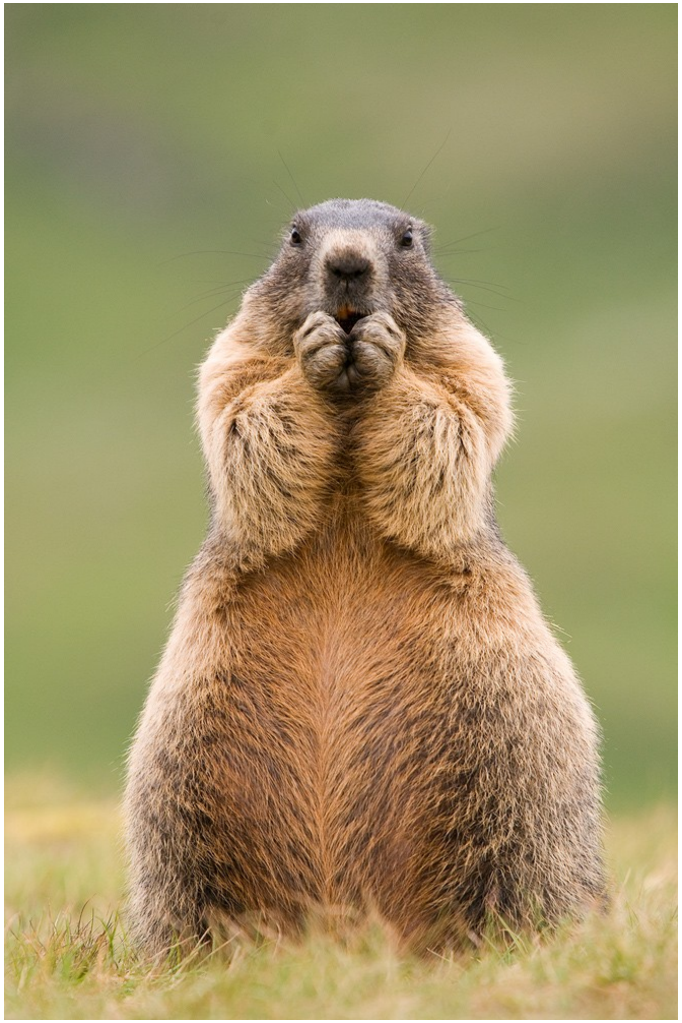
26. Oktober 2023

Wissenschafliches Messen

Was macht wissenschafltiches Messen aus?

Wissenschafliches Messen

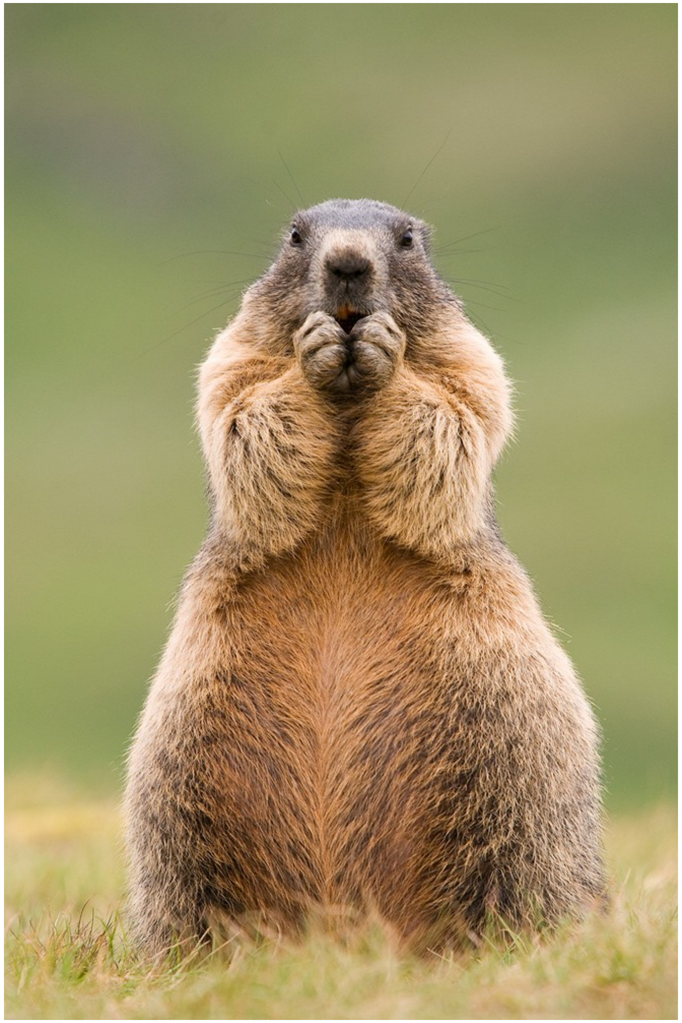
Was macht wissenschafltiches Messen aus?



Priv.-Doz. Dr. Roger Wolf
http://ekpwww.physik.uni-karlsruhe.de/~rwolf/

Wissenschafliches Messen

Was macht wissenschaftiches Messen aus?

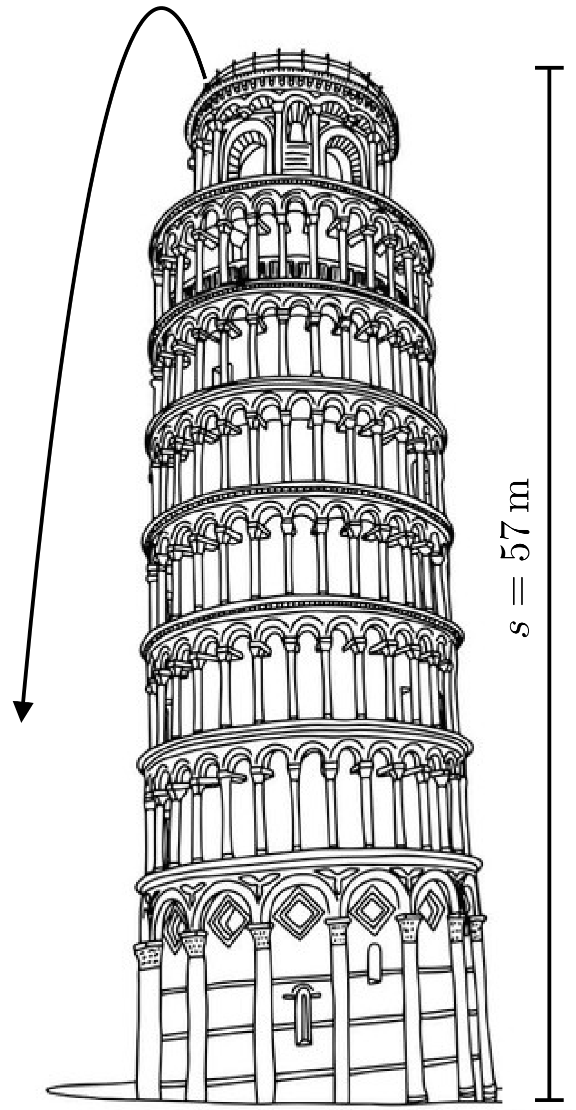


… Reproduzierbarkeit

Messung der Ergbeschleunigung g

$$\ell = \frac{1}{2} g\, t^2 \qquad g = \frac{2\ell}{t^2}$$

$$g(\ell, t) = \frac{2 \cdot 57\,\mathrm{m}}{(3,41\,\mathrm{s})^2} = 9,80\,\mathrm{m/s^2}$$



Messung der Ergbeschleunigung g

$$\ell = \frac{1}{2} g\, t^2 \qquad g = \frac{2\ell}{t^2}$$

$$g(\ell, t) = \frac{2 \cdot 57\,\mathrm{m}}{(3,41\,\mathrm{s})^2} = 9,80\,\mathrm{m/s}^2$$

- **Modellparameter bestimmt aus Beobachtung von t**

Messung der Ergbeschleunigung g

$$\ell = \frac{1}{2} g\, t^2 \qquad g = \frac{2\ell}{t^2}$$

$$g(\ell, t) = \frac{2 \cdot 57\,\mathrm{m}}{(3,41\,\mathrm{s})^2} = 9,80\,\mathrm{m/s^2}$$

- **Modellparameter bestimmt aus Beobachtung von** t
- Was ist das Modell?

Messung der Ergbeschleunigung g

$$\ell = \frac{1}{2} g \, t^2 \qquad g = \frac{2\ell}{t^2}$$

$$g(\ell, t) = \frac{2 \cdot 57\,\mathrm{m}}{(3,41\,\mathrm{s})^2} = 9,80\,\mathrm{m/s^2}$$

- **Modellparameter bestimmt aus Beobachtung von t**
- Was ist das Modell?
- Wozu überhaupt ein Modell?

Messung der Ergbeschleunigung g

$$\ell = \frac{1}{2} g\, t^2 \qquad g = \frac{2\ell}{t^2}$$

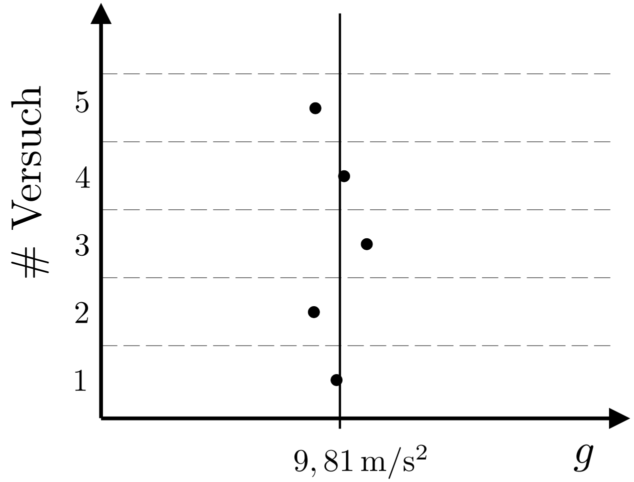
$$g(\ell, t) = \frac{2 \cdot 57\,\mathrm{m}}{(3,41\,\mathrm{s})^2} = 9,80\,\mathrm{m/s^2}$$

- **Modellparameter bestimmt aus Beobachtung von t**
- Was ist das Modell?
- Wozu überhaupt ein Modell?
- Messung reproduzierbar?

Messung der Ergbeschleunigung g

$$\ell = \frac{1}{2} g\, t^2 \qquad g = \frac{2\ell}{t^2}$$

$$g(\ell, t) = \frac{2 \cdot 57\,\mathrm{m}}{(3,41\,\mathrm{s})^2} = 9,80\,\mathrm{m/s}^2$$

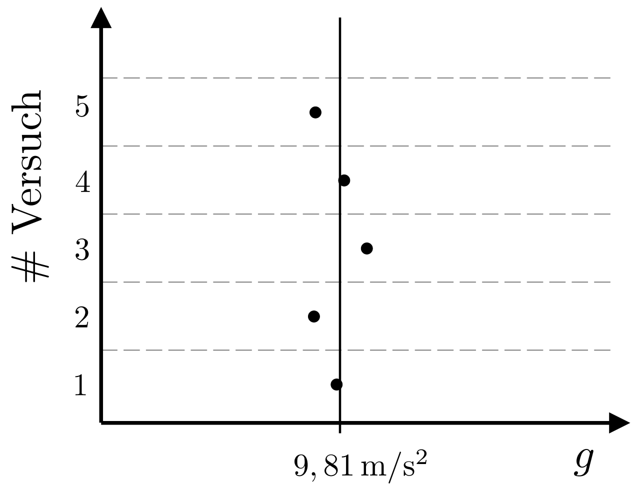


- **Modellparameter bestimmt aus Beobachtung von t**
- Was ist das Modell?
- Wozu überhaupt ein Modell?
- Messung reproduzierbar?

Messung der Ergbeschleunigung g

$$\ell = \frac{1}{2} g\, t^2 \qquad g = \frac{2\ell}{t^2}$$

$$g(\ell, t) = \frac{2 \cdot 57\,\mathrm{m}}{(3,41\,\mathrm{s})^2} = 9,80\,\mathrm{m/s^2}$$



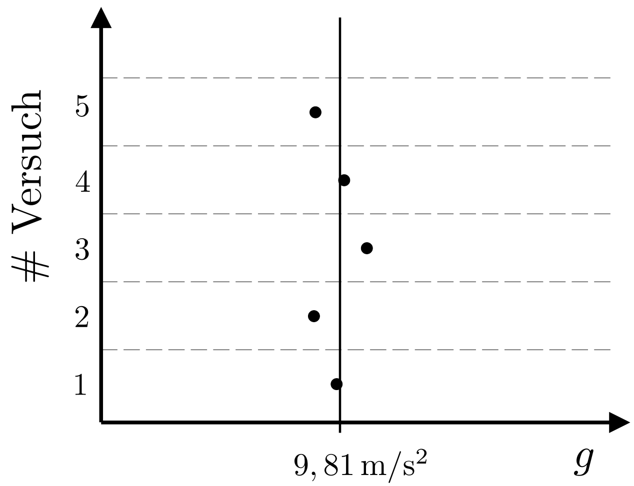
- **Modellparameter bestimmt aus Beobachtung von t**
- Was ist das Modell?
- Wozu überhaupt ein Modell?
- Messung reproduzierbar?

A posteriori: Ergebnis jeder einzelnen Messung historische Gewissheit

Messung der Ergbeschleunigung g

$$\ell = \frac{1}{2} g\, t^2 \qquad g = \frac{2\ell}{t^2}$$

$$g(\ell, t) = \frac{2 \cdot 57\,\mathrm{m}}{(3,41\,\mathrm{s})^2} = 9,80\,\mathrm{m/s^2}$$



- **Modellparameter bestimmt aus Beobachtung von t**
- Was ist das Modell?
- Wozu überhaupt ein Modell?
- Messung reproduzierbar?

A priori: Ausgang des Experiments ungewiss → Zufallsexperiment

Priv.-Doz. Dr. Roger Wolf
http://ekpwww.physik.uni-karlsruhe.de/~rwolf/

Reproduzierbarkeit

Wissenschaftliches Experiment → nur innerhalb bestimmter Grenzen reproduzierbar (→ Unsicherheiten)

Reproduzierbarkeit

Wissenschaftliches Experiment → nur innerhalb bestimmter Grenzen reproduzierbar (→ Unsicherheiten)

Konfidenz- oder Vertrauensintervall

Reproduzierbarkeit

Wissenschaftliches Experiment → nur innerhalb bestimmter Grenzen reproduzierbar (→ Unsicherheiten)

Konfidenz- oder Vertrauensintervall

Ohne zusätzliche Angabe eines Vertrauensintervalls ist jede wissenschaftliche Messung

Reproduzierbarkeit

Wissenschaftliches Experiment → nur innerhalb bestimmter Grenzen reproduzierbar (→ Unsicherheiten)

Konfidenz- oder Vertrauensintervall

Ohne zusätzliche Angabe eines Vertrauensintervalls ist jede wissenschaftliche Messung W E R T L O S

Vertrauensintervall

- **Wie kann ich zu einem solchen Vertrauensintervall gelangen?**
  - Autorität anderer (→ Datenblatt)
  - Eigene Einschätzung (→ Wiss. Intuition/Integrität)
  - Stichprobe (→ Mathematik)
  - Fortpflanzung aus bekannten Unsicherheiten (→ Modell)

Autorität anderer

METRA HIT 22 … 26S/M
Analog-Digital-Multimeter mit Signalgenerator

3-349-026-01
8/11.04

METRA HIT 22/23/24/25/26
- Präzisionsmultimeter (V, dB, Ω, F, Hz, °C/°F, V-▸|-)
- Auflösung: 10 µV, 10 mΩ
- Integrierte Quarzuhr für echtzeitbezogene MIN-/MAX-Registrierung
- Rechtecksignalgeneratorfunktionen
- Infrarot-Datenschnittstelle
- DKD-Kalibrierschein

METRA HIT 23/24/25/26
- Strommessung (10 A) direkt oder über Zangenstromwandler: ein Übersetzungsverhältnis von 1000:1 oder 10000:1 wird in der Anzeige berücksichtigt
- METRA HIT 23S: 16 A-Messbereich (ungesichert) speziell für Messungen an Stromwandlern

METRA HIT 22M/26M
- Großer Messdatenspeicher für bis zu 100000 Messwerten
- Echtzeitbezogener, quarzgesteuerter Datenlogger

METRA HIT 25/26
- Echteffektivwertmessung TRMS

Signalgeneratorfunktion

CAT IV

DKD

QUALITÄTSMANAGEMENTSYSTEM
DQS-zertifiziert nach DIN EN ISO 9001:2000 Reg.-Nr.1262

Referenzbedingungen

Umgebungstemperatur	+23 °C ±2 K
Relative Feuchte	40 ... 60%
Frequenz der Messgröße	45 ... 65 Hz
Kurvenform der Messgröße	Sinus
Batteriespannung	3 V ±0,1 V

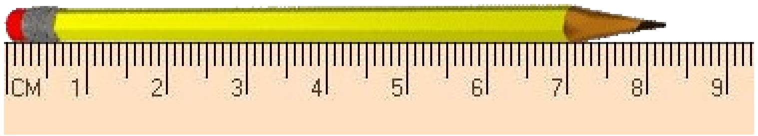
Technische Kennwerte

Mess-funktion	Messbereich	Auflösung bei Messbereichsendwert			Eingangsimpedanz		Eigenabweichung der höchsten Auflösung bei Referenzbedingungen		Überlastbarkeit 7)		Messrate		
							±(... % v. M. + ... D)	±(... % v. M. + ... D)					
		30000 1)	3000		—	~ / ≂	—	~ / ≂ 8)	Wert	Zeit	—	≂	~
V 4)	300 mV	10 µV			> 20 MΩ	5 MΩ // < 50 pF	0,05 + 3 7)	0,5 + 30 (> 300 D)	1000 V DC AC eff Sinus	dauernd	50 ms (22M/ 26M: 1 ms)	0,5 s	1 s
	3 V	100 µV			11 MΩ	5 MΩ // < 50 pF	0,05 + 3	0,2 + 30 (> 300 D)					
	30 V	1 mV			10 MΩ	5 MΩ // < 50 pF	0,05 + 3	0,2 + 30 (> 300 D)					
	300 V	10 mV			10 MΩ	5 MΩ // < 50 pF	0,05 + 3	0,2 + 30 (> 300 D)					
	1000 V	100 mV			10 MΩ	5 MΩ // < 50 pF	0,05 + 3	0,2 + 30 (> 300 D)					

1) Anzeige: 4¾ Stellen; für die Speicherung und Übertragung von Messwerten ist eine andere Auflösung und Abtastrate einstellbar im Menü rAtE.
2) Stoppuhr: Format: mm:ss:h mit m=Minute, s=Sekunde und h=Hundertstelsekunde, max.: 99:59.9; nur über Tasten bedienbar
3) niedrigste messbare Frequenz bei sinusförmigem Messsignal symmetrisch zum Nullpunkt
4) METRA HIT 26S/M und 25S: Echte Effektivwertmessung TRMS
5) ohne 16 A-Sicherung
6) Anzeige bis max. 1,8 V, darüber Überlauf „OL".
7) bei 0 ° ... + 40 °C
8) Werte < 100 Digit werden unterdrückt
15 (20) ... 45 ... 65 Hz ... 20 (1) kHz Sinus. Einflüsse siehe Seite 4.
9) 12 A – 5 min, 16 A – 30 s, METRA HIT23S: 16 A 10 min.
10) bei Funktion „Nullpunkteinstellung" aktiv, Anzeige ZERO
11) die Amplitude der Eingangsspannung darf folgende Werte nicht unterschreiten/überschreiten:

Eigene Einschätzung

- Bona fides, **gesunder Menschenverstand**
- Anhaltspunkte:

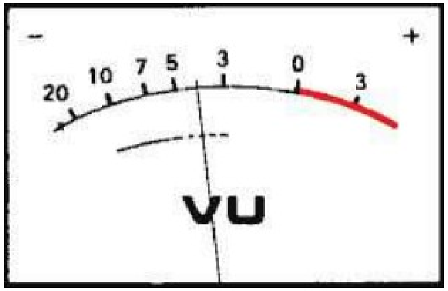


$$L = (82 \pm 0,5)\,\mathrm{mm}$$

$$U = (13,06 \pm 0,005)\,\mathrm{V}$$

$\oplus$ Angaben aus Datenblatt

$\oplus$ Anzeigefluktuationen

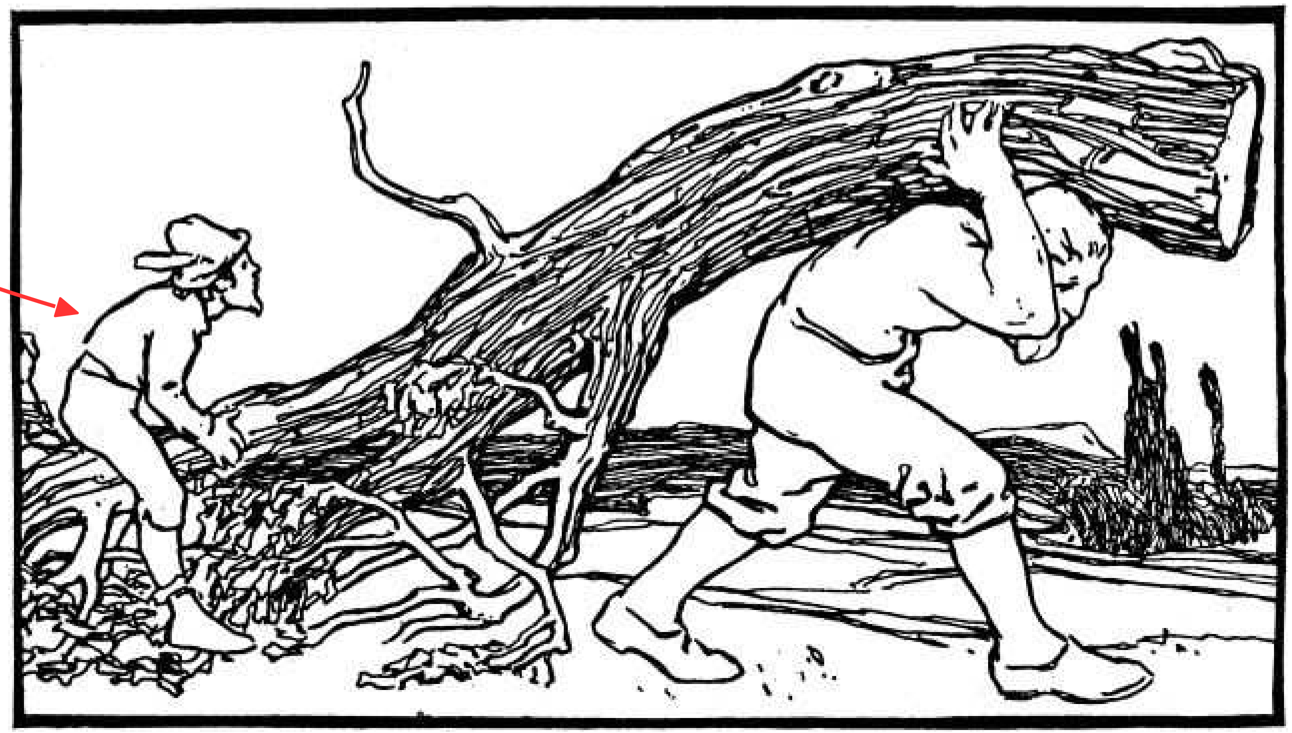


$$U = (4,0 \pm 0,2)\,\mathrm{V}$$

Schmaler Grat

Stichprobe

Let’s maths do the work



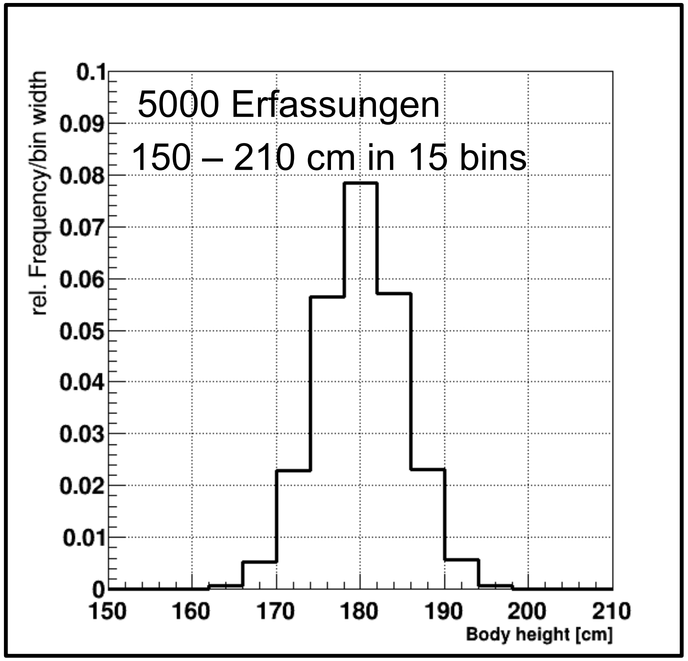
Grundbegriffe der Statistik

- Erfassung Körpergröße x von 5000 Einwohnern (über 18 Jahre) in Karlsruhe
- Menge aller Einwohner in Karlsruhe $\rightarrow$ **Grundgesamtheit**
- Messung mit 5000 Probanden $\rightarrow$ **Stichprobe** (engl. *sample*)
- $x \rightarrow$ (kontinuierlich verteilte) **Zufallsvariable**, Ergebnis eines **Zufallsexperiments**

Karlsruhe: Das Fest

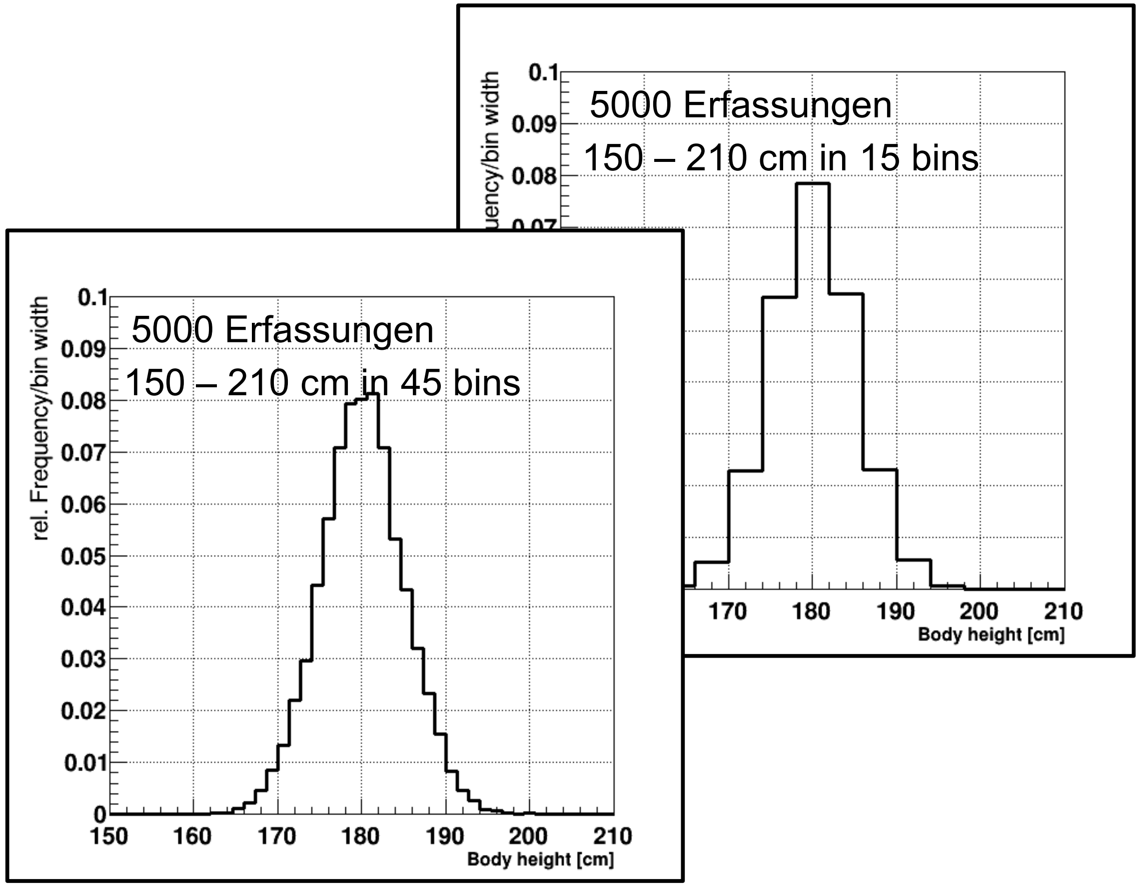
Histogramm

- **Darstellung** durch Histogramm, dividiert durch Stichprobenlänge (=5000) und Binbreite.



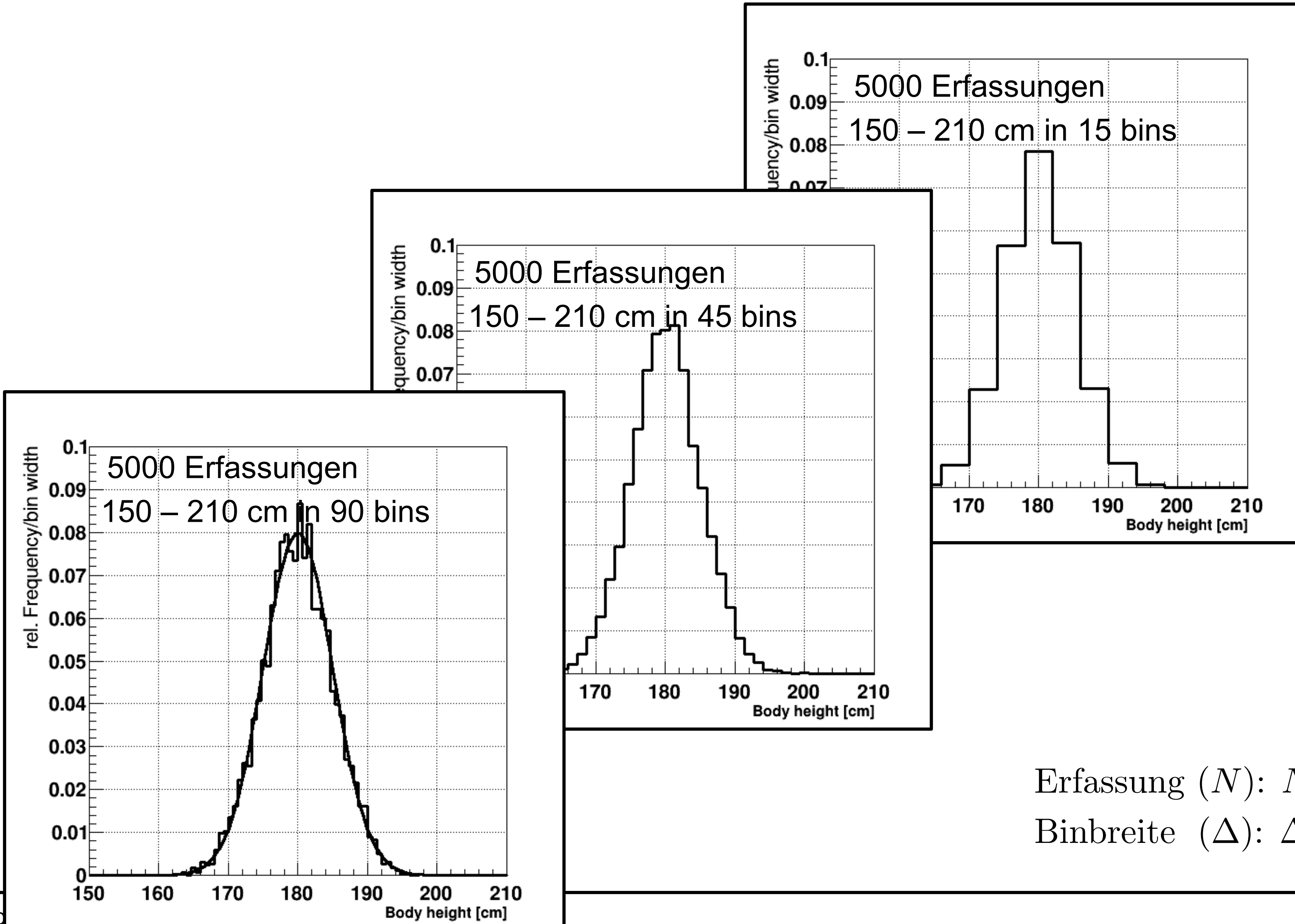
Histogramm

- **Darstellung** durch Histogramm, dividiert durch Stichprobenlänge (=5000) und Binbreite.



Histogramm

- **Darstellung** durch Histogramm, dividiert durch Stichprobenlänge (=5000) und Binbreite.



$$\text{Erfassung } (N): \; N \to \infty$$
$$\text{Binbreite } (\Delta): \; \Delta \to 0$$

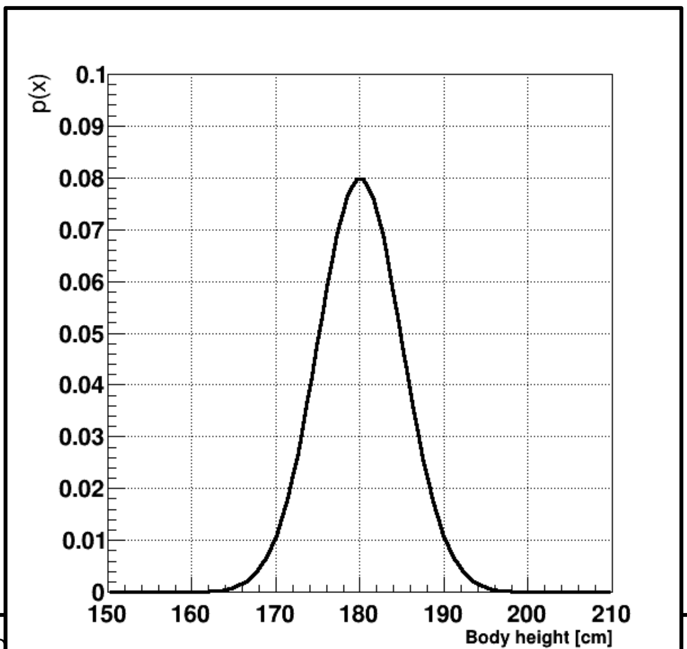
Wahrscheinlichkeitsdichte

Ist x eine kontinulierlich verteilte Zufallsvariable und die Wahrscheinlichkeitsverteilung $P(x)$ über dem Ergebnisraum Ω stetig differenzierbar, dann bezeichnen wir

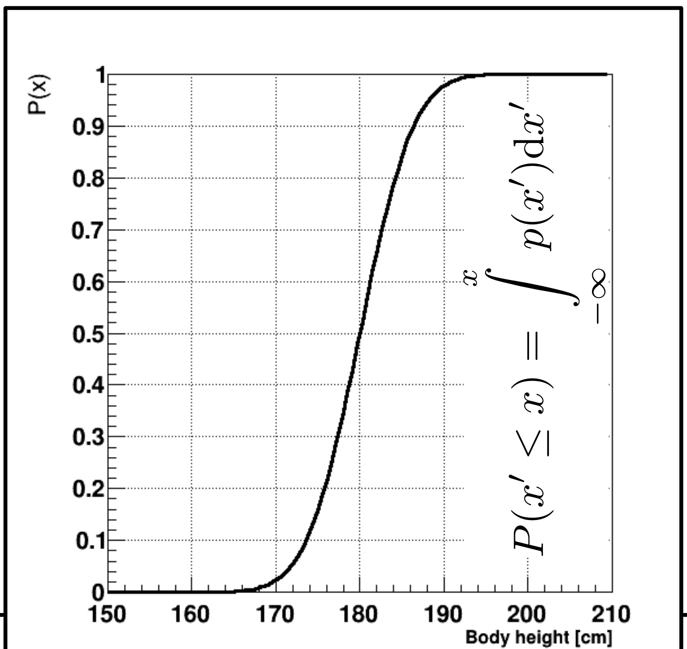
$$p(x) = \frac{\mathrm{d}P}{\mathrm{d}x}(x)$$

als Wahrscheinlichkeitsdichtefunktion von x.

Wahrscheinlichkeitsdichte (engl. *probability density function*, PDF)



Kumulative Wahrscheinlichkeitsfunktion (engl. *cumulative distribution function*, CDF)



Erwartungswert

Ist x eine kontinuierlich (diskret) verteilte Zufallsvariable und $p(x)$ $(P(x))$ die zugehörige Wahrscheinlichkeitsdichte(-verteilung) über Ω. Dann bezeichnet man die Größe

$$E[x] = \int\limits_{\Omega} x\, p(x) \mathrm{d}x \qquad \text{(kontinuierlich)}$$

$$E[x] = \sum_{\Omega} x_i\, P(x_i) \qquad \text{(diskret)}$$

als den Erwartungswert für x über Ω.

- Für eine vorgegebene Wahrscheinlichkeitsdichte(-verteilung) ist $E[x]$ eine <u>Zahl</u> und keine Funktion von x (andere gängige Bezeichnungen: μ, $\langle x \rangle$).
- Der Erwartungswert ist linear in x:

$$E\,[\,\alpha\, x + \beta\, y\,] = \alpha\, E[\,x\,] + \beta\, E[\,y\,]$$

Varianz

Man bezeichnet $E\left[(x-x_0)^n\right]$ als das n-te algebraische Moment um x_0. In der Statistik sind die folgenden Spezialfälle für $x_0 = E[x]$ von Relevanz:

$$\text{0-tes Moment:} \quad E[(x-E[x])^0] = \int\limits_\Omega p(x)\,\mathrm{d}x = 1$$

$$\text{1-tes Moment:} \quad E[(x-E[x])^1] = \int\limits_\Omega (x-E[x])\,p(x)\,\mathrm{d}x = 0$$

$$\begin{aligned}\text{2-tes Moment:} \quad E[(x-E[x])^2] &= \int\limits_\Omega (x-E[x])^2\,p(x)\,\mathrm{d}x \\ &= \underbrace{\int\limits_\Omega x^2\,p(x)\,\mathrm{d}x}_{E[x^2]} - \underbrace{2E[x]\int\limits_\Omega x\,p(x)\,\mathrm{d}x}_{E[x]^2} + \underbrace{E[x]^2\int\limits_\Omega p(x)\,\mathrm{d}x}_{E[x]^2}\end{aligned}$$

Das 2-te algebraische Moment um $E[x]$

$$\mathrm{var}[x] = E\left[(x-E[x])^2\right] = E[x^2] - E[x]^2$$

heißt Varianz von x über Ω, $\sigma_x = \sqrt{\mathrm{var}[x]}$ heißt Standardabweichung.

Priv.-Doz. Dr. Roger Wolf
http://ekpwww.physik.uni-karlsruhe.de/~rwolf/

Wahrscheinlichkeit

- Wahrscheinlichkeit für $x = (180 \pm 1)\mathrm{cm}$ nicht aus $p(x)|_{x=180}$ sondern aus:

$$\int\limits_{179\,\mathrm{cm}}^{181\,\mathrm{cm}} p(x')\mathrm{d}x' = P(x \leq 181\,\mathrm{cm}) - P(x \leq 179\,\mathrm{cm})$$

Wahrscheinlichkeit

- Wahrscheinlichkeit für $x = (180 \pm 1)\mathrm{cm}$ nicht aus $p(x)|_{x=180}$ sondern aus:

$$\int\limits_{179\,\mathrm{cm}}^{181\,\mathrm{cm}} p(x')\mathrm{d}x' = P(x \leq 181\,\mathrm{cm}) - P(x \leq 179\,\mathrm{cm})$$

$p(x)$ auf Grundgesamtheit nicht bekannt!

Wahrscheinlichkeit

- Wahrscheinlichkeit für $x = (180 \pm 1)\mathrm{cm}$ nicht aus $p(x)|_{x=180}$ sondern aus:

$$\int_{179\,\mathrm{cm}}^{181\,\mathrm{cm}} p(x')\mathrm{d}x' = P(x \leq 181\,\mathrm{cm}) - P(x \leq 179\,\mathrm{cm})$$

$E[x]$ und $\mathrm{var}[x]$ aus Stichprobe zielsicher abschätzbar**, selbst ohne Kenntnis von $p(x)$!**

Mittelwert der Stichprobe

Die Größe

$$\overline{x} = \frac{1}{n}\sum_{i\leq n} x_i$$

heißt Mittelwert der Stichprode. Ihr Erwartungswert ($E[\overline{x}]$) und ihre Varianz ($\mathrm{var}[\overline{x}]$) sind:

$$E[\overline{x}] = E\left[\frac{1}{n}\sum_{i\leq n} x_i\right] = \frac{1}{n}\sum_{i\leq n}\underbrace{E[x_i]}_{\equiv\,\mu} = \frac{1}{n}\sum_{i\leq n}\mu = \mu$$

$$\mathrm{var}[\overline{x}] = E[\overline{x}^2] - E[\overline{x}]^2 = E\left[\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)\left(\frac{1}{n}\sum_{j=1}^{n} x_j\right)\right] - \mu^2 = \frac{1}{n^2}\sum_{i,j=1}^{n} E[x_i x_j] - \mu^2$$

$$= \frac{1}{n^2}\left[(n^2-n)\mu^2 + n(\mu^2+\sigma^2)\right] - \mu^2 = \sigma^2/n\,.$$

Dabei sind μ und σ^2 der Erwartungswert und die Varianz der Einzelmessung und n die Länge der Stichprobe.

Varianz der Stichprobe

Die Größe

$$s^2 = \frac{1}{n-1}\sum_{i\leq n}(x_i - \overline{x})^2 = \frac{n}{n-1}\left(\overline{x^2} - \overline{x}^2\right)$$

heißt Varianz der Stichprobe. Ihr Erwartungswert ($E[s^2]$) und ihre Varianz sind:

$$E[s^2] = \sigma^2$$

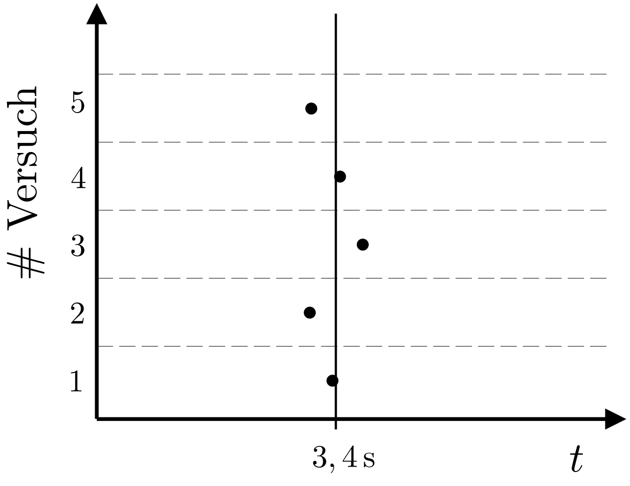
$$\mathrm{var}[s^2] = \frac{1}{n}\left(\mu_4 - \frac{n-3}{n-4}\mu_2^2\right),$$

wobei μ_k das k-te zentrale Moment um μ, und μ und σ^2 der Erwartungswert und die Varianz der Einzelmessung sind.

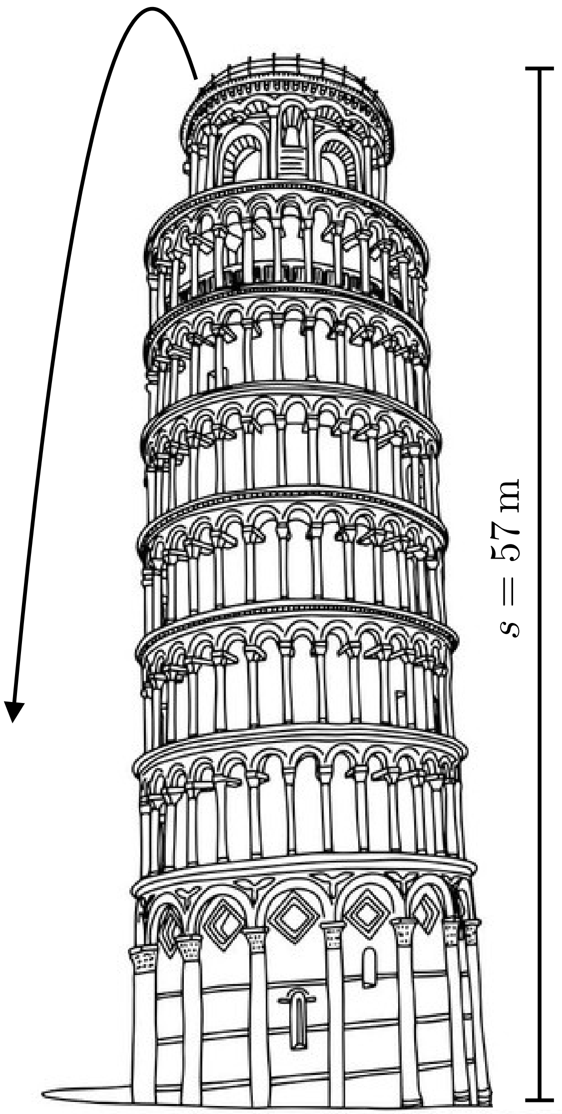
- Schätze μ_k durch $m_k = \dfrac{1}{n-1}\sum\limits_{i\leq n}(x_i - \overline{x})^k$ ab
- Beachte Normierung auf $n-1$ statt n (**Besselkorrektur**)

Statistik $\leftrightarrow$ Wir

- N (unabhängige!) Messungen t_i
- Berechne $\bar{t}$
- Berechne s_t^2
- $E[\,\bar{t}\,] = \mu_t; \quad \hat{\mathrm{var}}[\,\bar{t}\,] = \dfrac{s_t^2}{N}$

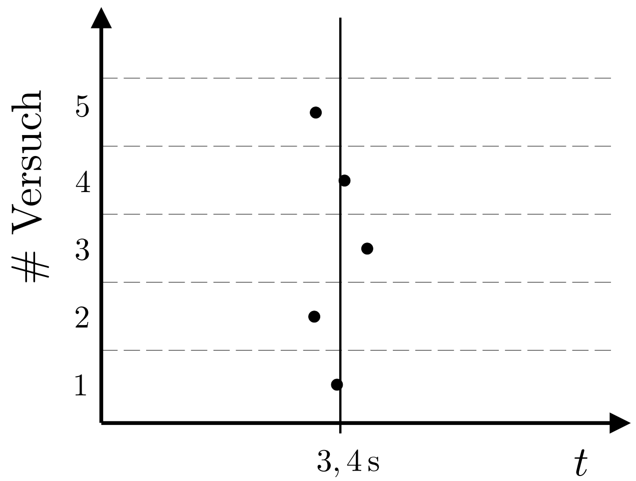


$$\bar{t} = (3,405 \pm 0,07)\,\mathrm{s}$$



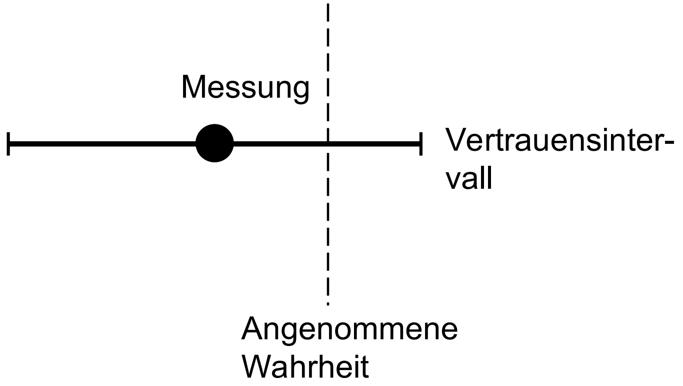
Vertrauensintervall?

- N (unabhängige!) Messungen t_i
- Berechne $\bar{t}$
- Berechne s_t^2
- $E[\,\bar{t}\,] = \mu_t; \quad \hat{\mathrm{var}}[\,\bar{t}\,] = \dfrac{s_t^2}{N}$



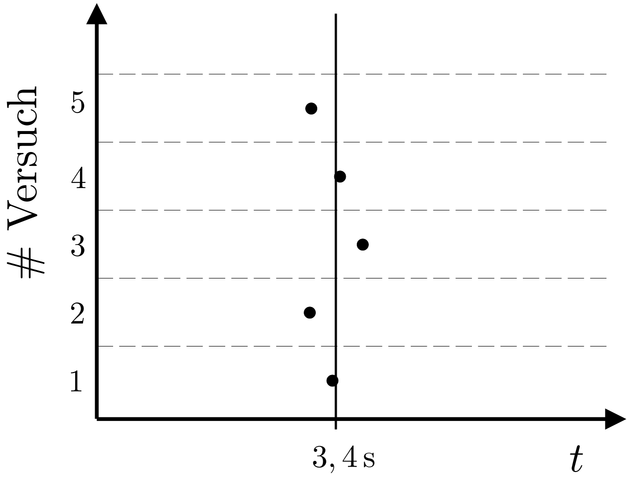
$$\bar{t} = (3,405 \pm 0,07)\,\mathrm{s}$$

- **Konventionelle (Modell)Annahme:** Zufallsvariablen normalverteilt
- Unsicherheit als $\pm 1\,\sigma$ Vertrauensintervall
- Vertrauensintervall beinhaltet (unbekannte) Wahrheit mit $P = 0,68$ (→ Repräsentationsschluss)



Kompatibilität?

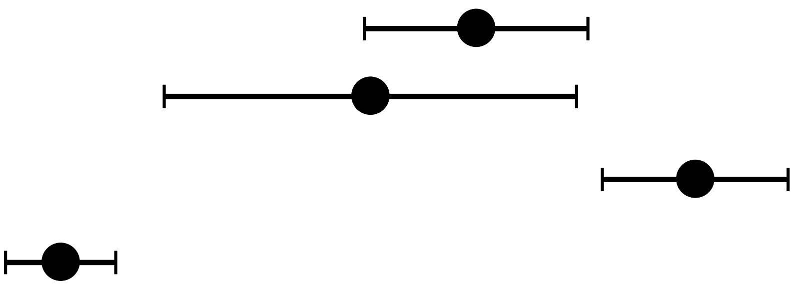
- N (unabhängige!) Messungen t_i
- Berechne $\bar{t}$
- Berechne s_t^2
- $E[\,\bar{t}\,] = \mu_t; \quad \hat{\text{var}}[\,\bar{t}\,] = \frac{s_t^2}{N}$



$$\bar{t} = (3,405 \pm 0,07)\,\text{s}$$

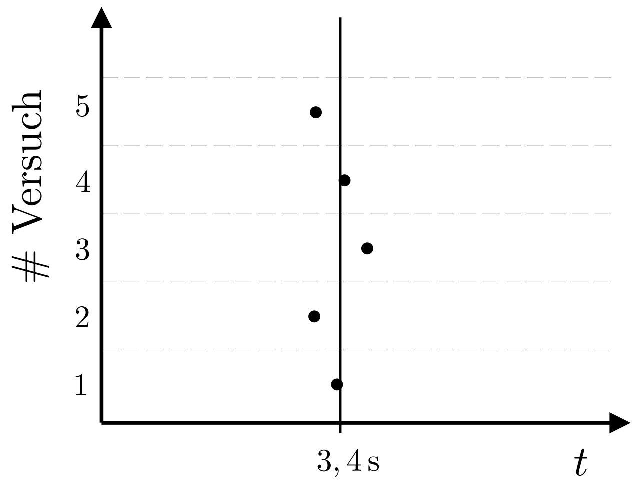
- **Zwei Messungen inkompatibel:**

$$\delta = \frac{t_i - t_j}{\Delta(t_i - t_j)} \gg 1 \qquad \text{(Pull)}$$



Nachkommastellen

- N (unabhängige!) Messungen t_i
- Berechne $\bar{t}$
- Berechne s_t^2
- $E[\,\bar{t}\,] = \mu_t; \quad \hat{\mathrm{var}}[\,\bar{t}\,] = \dfrac{s_t^2}{N}$



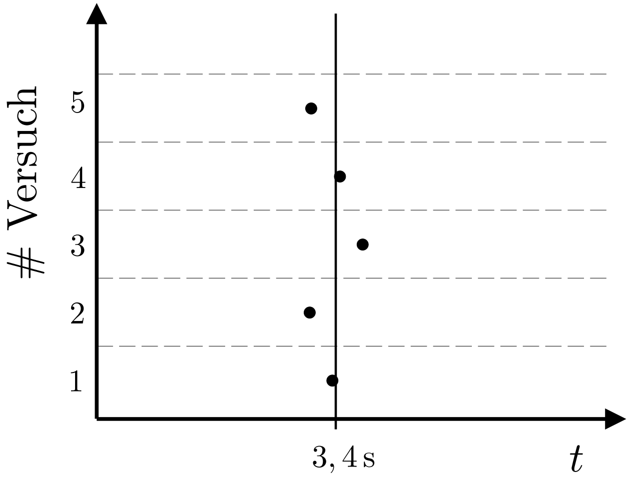
$$\bar{t} = (3,405 \pm 0,07)\,\mathrm{s}$$

- **Jenseits der bestimmten Unsicherheit $\rightarrow$ nutzlos**
- Faustregel: Erste Stelle, ab der $\Delta t \neq 0$
- Ausnahme: +1 Stelle bei Weiterverar-beitung

Fortpflanzung aus bekannten Unsicherheiten

$$\ell = \frac{1}{2} g\, t^2 \qquad g = \frac{2\ell}{t^2}$$

$$g(\ell, \bar{t}) = \frac{2 \cdot 57\,\mathrm{m}}{(3,405\,\mathrm{s})^2} = (9,83 \pm 0,20)\,\mathrm{m/s^2}$$



$$\bar{t} = (3,405 \pm 0,07)\,\mathrm{s}$$

Unsicherheiten:

- Funktionen von Zufallsvariablen sind selbst Zufallsvariablen
- Folgen Wahrscheinlichkeitsdichte
- **Erhaltung der Wahrscheinlichkeit:**

$$p(t)\,\mathrm{d}t = q(g)\,\mathrm{d}g$$

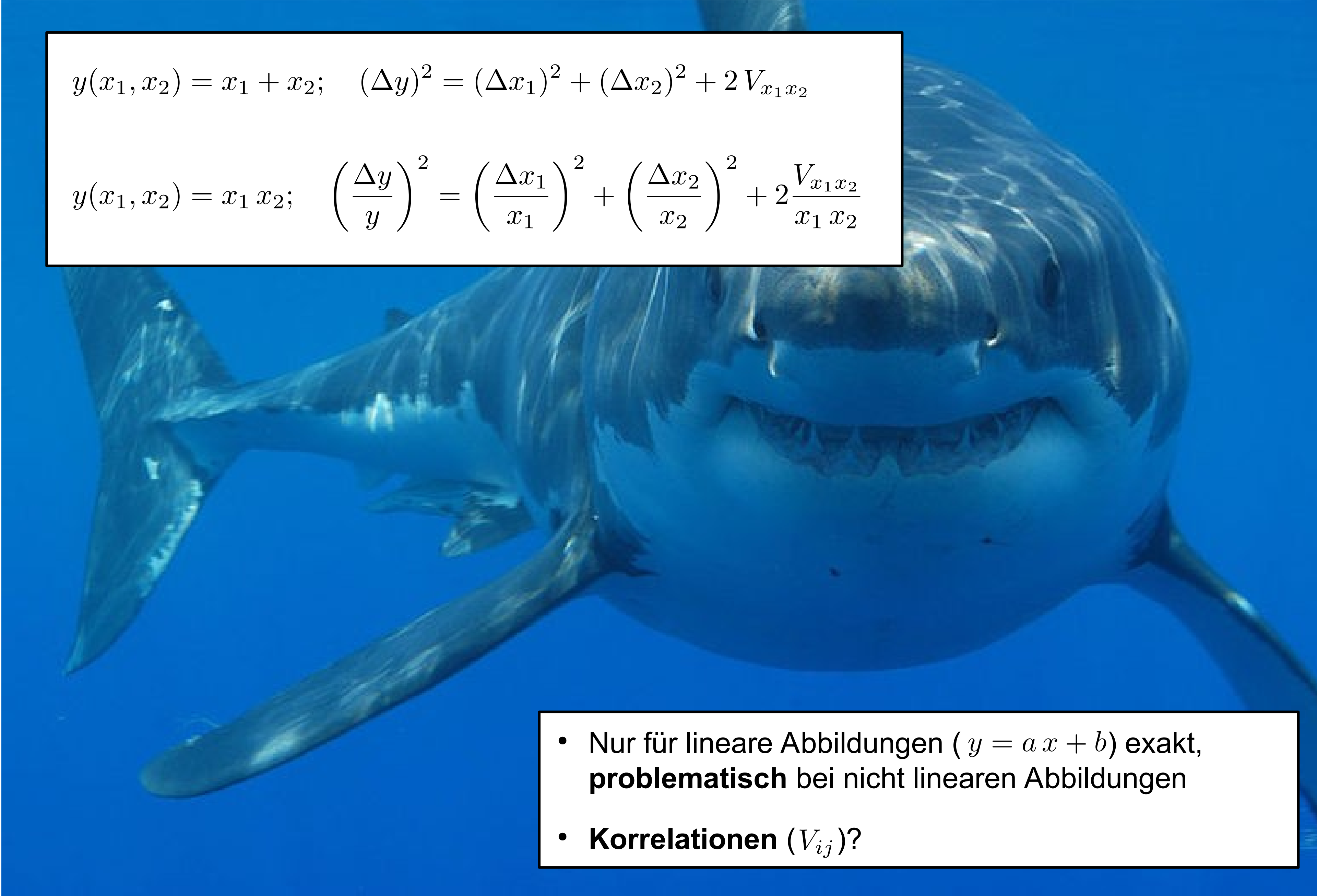
$$q(g) = p(t) \left| \frac{\partial g(t)}{\partial t} \right|$$

Beispiele

$$y(x_1, x_2) = x_1 + x_2; \quad (\Delta y)^2 = (\Delta x_1)^2 + (\Delta x_2)^2 + 2\,V_{x_1 x_2}$$

$$y(x_1, x_2) = x_1\,x_2; \quad \left(\frac{\Delta y}{y}\right)^2 = \left(\frac{\Delta x_1}{x_1}\right)^2 + \left(\frac{\Delta x_2}{x_2}\right)^2 + 2\frac{V_{x_1 x_2}}{x_1\,x_2}$$

Tücken



$$y(x_1, x_2) = x_1 + x_2; \quad (\Delta y)^2 = (\Delta x_1)^2 + (\Delta x_2)^2 + 2\,V_{x_1 x_2}$$

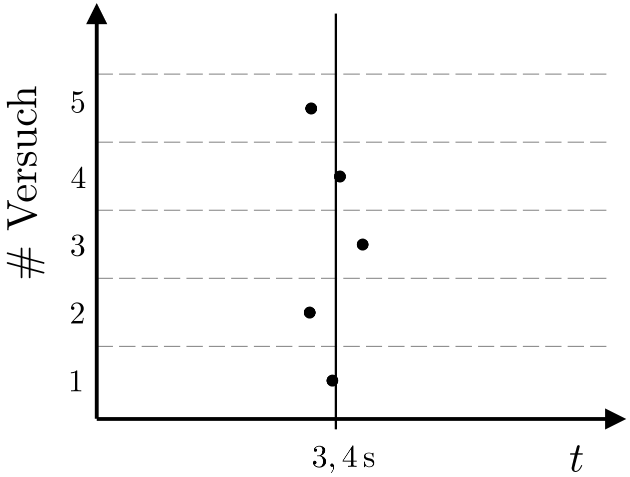
$$y(x_1, x_2) = x_1\,x_2; \quad \left(\frac{\Delta y}{y}\right)^2 = \left(\frac{\Delta x_1}{x_1}\right)^2 + \left(\frac{\Delta x_2}{x_2}\right)^2 + 2\frac{V_{x_1 x_2}}{x_1\,x_2}$$

- Nur für lineare Abbildungen ($y = a\,x + b$) exakt, **problematisch** bei nicht linearen Abbildungen
- **Korrelationen** (V_{ij})?

Was vergessen?

$$\ell = \frac{1}{2} g\, t^2 \qquad g = \frac{2\ell}{t^2}$$

$$g(\ell, \bar{t}) = \frac{2 \cdot 57\,\mathrm{m}}{(3,405\,\mathrm{s})^2} = (9,83 \pm 0,20)\,\mathrm{m/s^2}$$

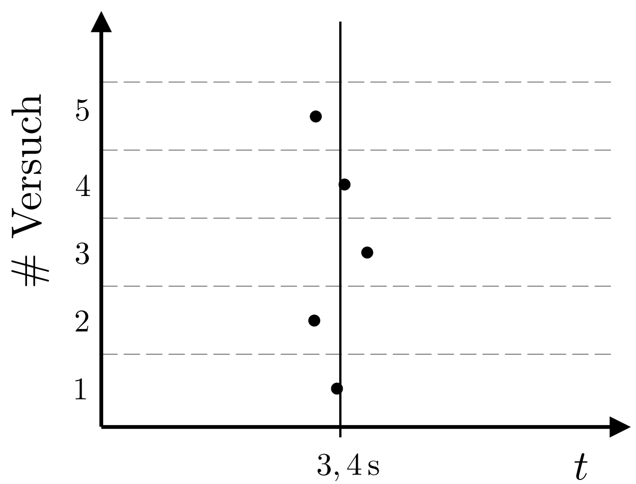


$$\bar{t} = (3,405 \pm 0,07)\,\mathrm{s}$$

Was vergessen!

$$\ell = \frac{1}{2} g\, t^2 \qquad g(\ell, t, \ldots)$$

$$g(\ell, \bar{t}) = \frac{2 \cdot 57\,\mathrm{m}}{(3,405\,\mathrm{s})^2} = (9,83 \pm 0,20)\,\mathrm{m/s}^2$$



$$\bar{t} = (3,405 \pm 0,07)\,\mathrm{s}$$

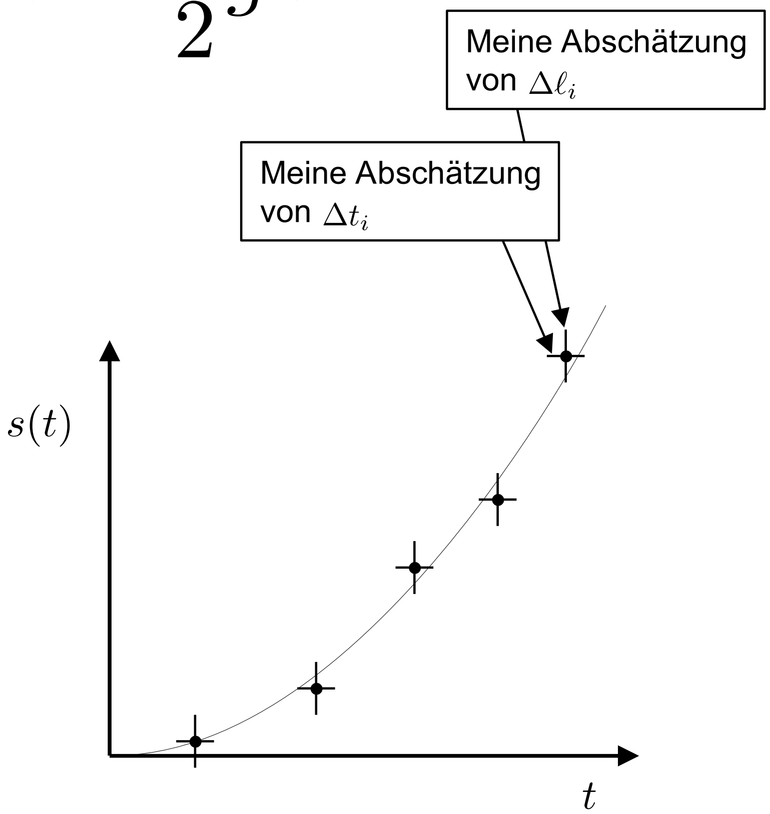
Außerdem:

- Höhe des Untergrunds?
- Anfangsgeschwindigkeit?
- Luftreibung?
- Auftrieb?
- Verborgene Parameter?
  $$g(r, M) = G\,\frac{M}{r^2}$$

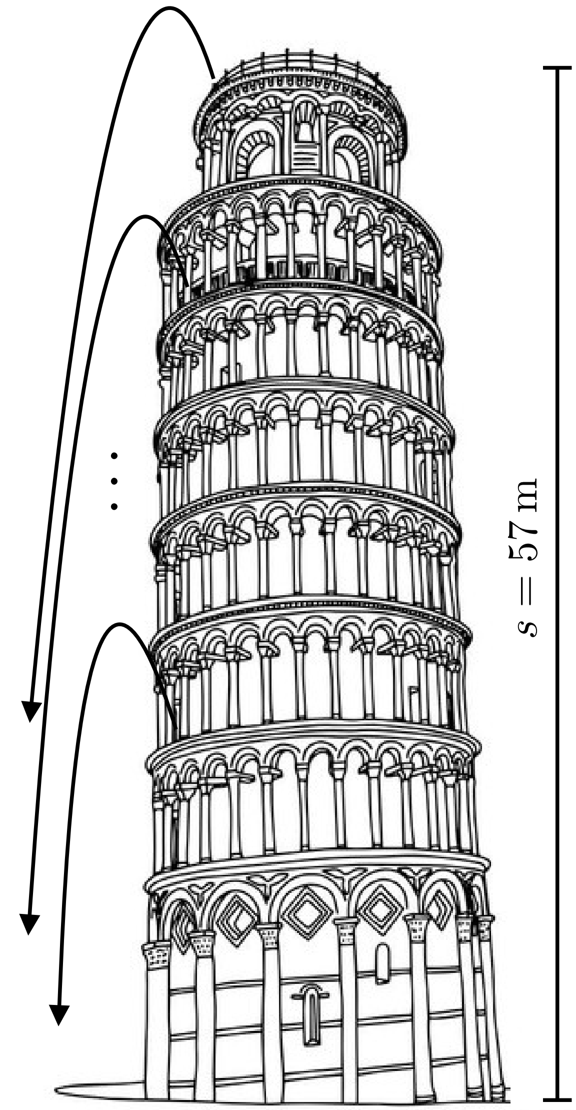
Priv.-Doz. Dr. Roger Wolf
http://ekpwww.physik.uni-karlsruhe.de/~rwolf/

Bessere Messung von g

$$\ell = \frac{1}{2} g\, t^2$$



$g(\{(\ell_i, t_i)\})$ hängt von $\{(\ell_i, t_i)\}$ ab!



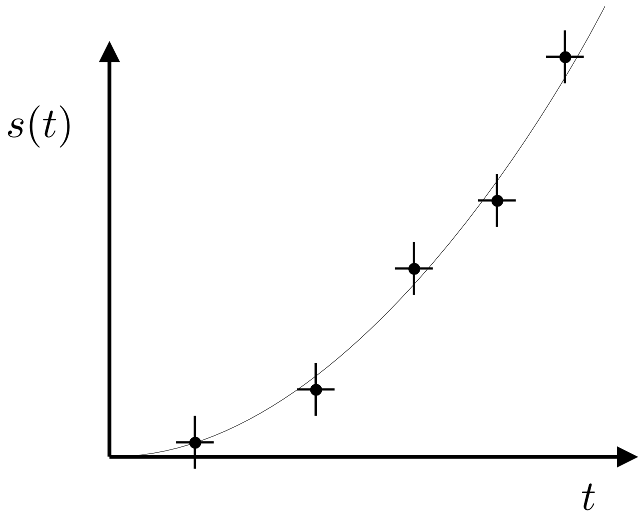
Neu $\leftrightarrow$ Alt

$$\ell = \frac{1}{2} g\, t^2$$

$$g = \frac{2\ell}{t^2}$$

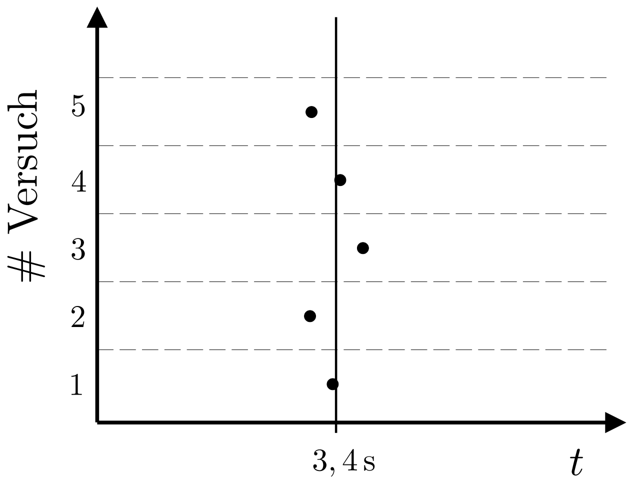
Neues Modell:

- Messe $\{(\ell_i, t_i)\}$ (stat. Unsicherheiten)



Altes Modell:

- Messe t (stat. Unsicherheit)
- ℓ extern bestimmter Parameter des Modells (syst. Unsicherheit)



Einordnung: Unsicherheiten

- **Statistische** Unsicherheiten:
  Alle Parameter, die ich selbst (in situ) (mit)gemessen habe
- **Systematische** Unsicherheiten:
  Alles was (als Annahme) von außen ins Modell eingeführt wurde
- Externe Parameter ohne Unsicherheiten? → modelabhängige Messung!

Goodness of fit

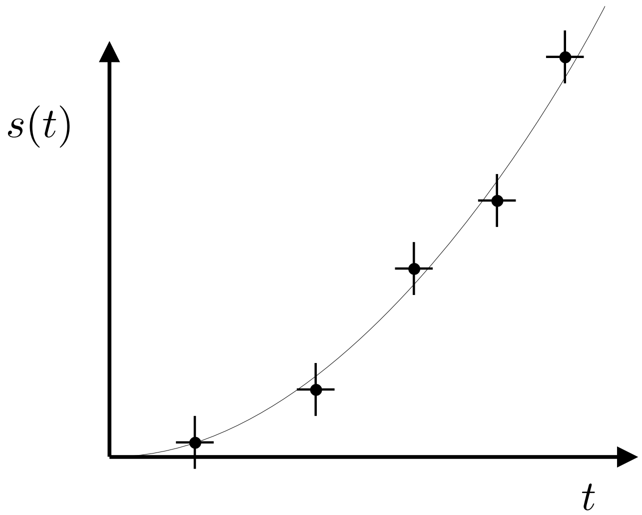
- **Q: Woher weiss ich, dass mein Modell stimmt?**

Goodness of fit

- **Q: Woher weiss ich, dass mein Modell stimmt? – A: Das weiss ich nicht.**

Goodness of fit

- **Q: Woher weiss ich, dass mein Modell stimmt? – A: Das weiss ich nicht.**
- Kompatibilität (→ Widerspruchsfreiheit) mit Daten



χ^2-Test:

- Abstand Modell-Messpunkt

$$d_i\left((\ell(g,\cdot),t(g,\cdot))-(\ell_i,t_i)\right)$$

$$z \equiv \sum_{i=1}^{n} \frac{d_i^2}{\Delta d_i}$$

- $d(g,\ell,t)$ normalverteilt? → z folgt $\boldsymbol{\chi^2(d,n-k)}$ **Verteilung**
- k Anzahl freier Parameter im Modell
- $n' \equiv n-k$ Anzahl **Freiheitsgrade**

χ^2-Verteilung

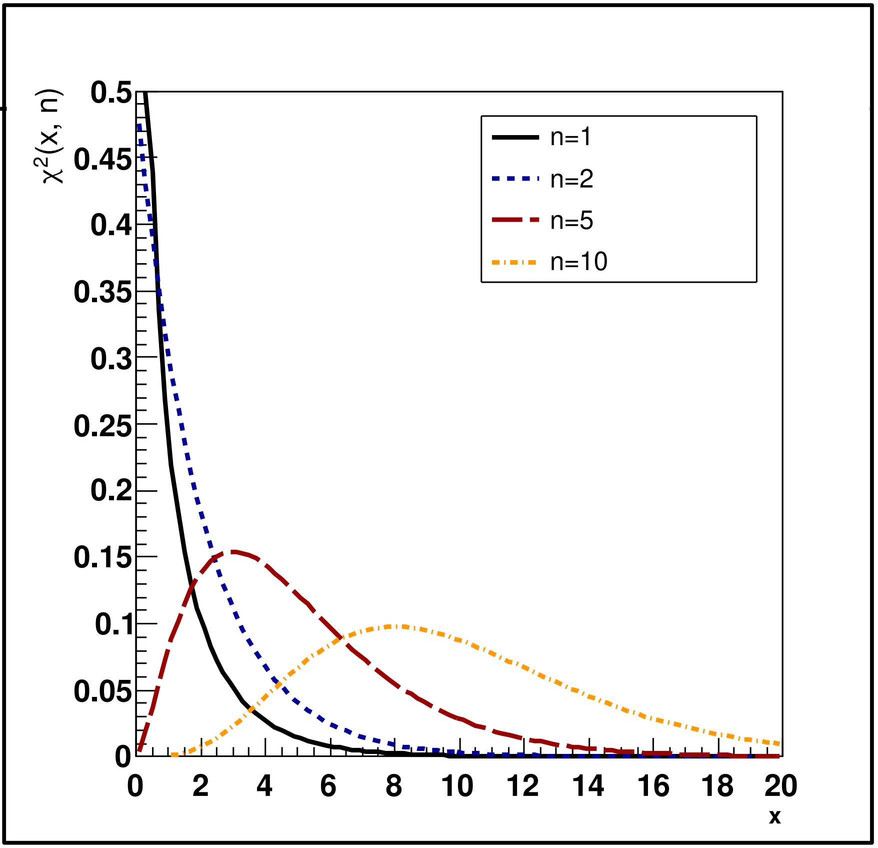
$$\chi^2(x,n') = \frac{1}{2^{n/2}\Gamma(n'/2)} x^{n'/2-1} e^{-x/2}$$

mit:

$$\Gamma(x) = \int e^{-t} t^{x-1} \mathrm{d}t$$

$$E[x] \quad = n' \qquad\qquad \text{(Erwartungswert)}$$

$$\mathrm{var}[x] = 2n' \qquad\qquad \text{(Varianz)}$$



χ^2-Test

- Model kann Daten beschreiben ☑
- $d(g, \ell, t)$ normalverteilt ☑

$$\rightarrow \; E[\,z\,] = n - k\,; \qquad \frac{E[\,z\,]}{n-k} = 1$$

$$\frac{z_{\mathrm{obs}}}{n-k} \ll 1 \qquad \Rightarrow \qquad \text{Fehler zu groß abgeschätzt oder zu viele Parameter im Modell?}$$

$$\frac{z_{\mathrm{obs}}}{n-k} \gg 1 \qquad \Rightarrow \qquad \text{Fehler zu klein abgeschätzt oder Modell inkompatibel.}$$

Kaffee?

- Hier diskutierte Methoden implementiert in kafe2

kafe2 2.8.1 SITE PAGE INSTALLING KAFE2 » SOURCE

kafe2 documentation

Welcome to **kafe2**, the *Karlsruhe Fit Environment 2*!

kafe2 is a data fitting framework originally designed for use in undergraduate physics lab courses. It provides a *Python* toolkit for fitting models to data via the maximum likelihood method as well as visualizing the fit results. A quick rundown of why you'd want to use *kafe2* can be found here. The gist of it is that *kafe2* provides a simple, user-friendly interface for state-of-the-art statistical methods. It relies on Python packages such as `numpy` and `matplotlib`, and can use either `scipy` or the minimizer *Minuit* contained in the Python package *iminuit* as the numerical optimization backend.

The first chapter of this documentation gives detailed installation instructions. The Beginner's Guide explains basic *kafe2* usage to cover simple cases (both Python code and kafe2go). The User Guide and the kafe2go Guide describe advanced *kafe2* use with Python code or *kafe2go*. The next chapter explains the mathematical foundations upon which *kafe* is built. While strictly speaking not required to use *kafe2*, reading the theory chapter is strongly recommended to understand which features to use in a state-of-the-art data analysis (regardless of whether *kafe2* or another data analysis tool is used). The Developer Guide covers topics that are only relevant if you want to work on *kafe2* as a developer (still very much WIP). Finally, the API Documentation provides a full description of the user-facing *kafe2* application programming interface.

PhyPraKit?

- Lightweight-Version PhyPraKit

PhyPraKit 1.2.6 documentation » **PhyPraKit Documentation**

Table of Contents

PhyPraKit Documentation

About

Version 1.2.6, Date 2023-10-14

PhyPraKit is a collection of python modules for data visualization and analysis in experimental laboratory courses in physics and is in use in the Department of Physics at Karlsruhe Institute of Technology (KIT). As the modules are intended primarily for use by undergraduate students in Germany, the documentation is partly in German language, in particular the description of the examples.

Created by:

- Guenter Quast <guenter (dot) quast (at) online (dot) de>

A pdf version of this documentation is available here: PhyPraKit.pdf.

Installation:

To use PhyPraKit, it is sufficient to place the directory *PhyPraKit* and all the files in it in the same directory as the python scripts importing it.

Installation via *pip* is also supported. After downloading, execute:

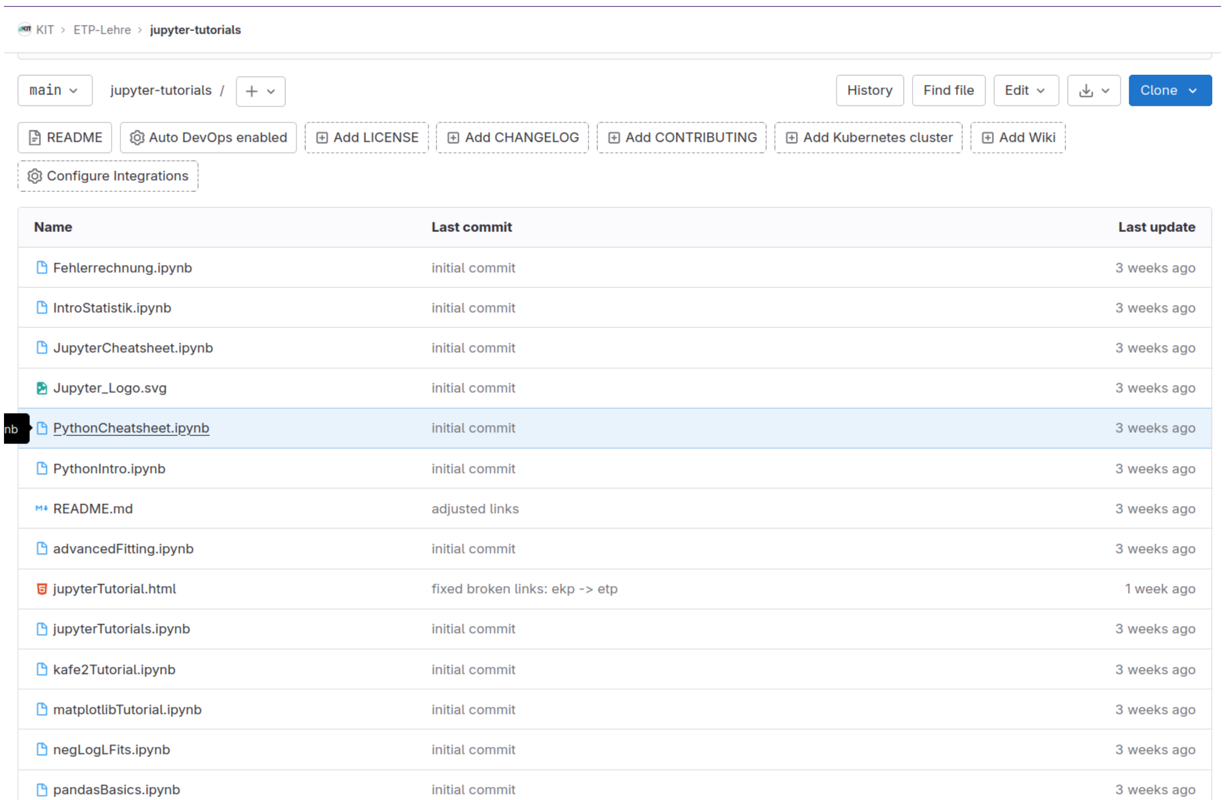
```
pip install --user .
```

in the main directory of the *PhyPraKit* package (where *setup.py* is located) to install in user space.

Comfortable installation via the PyPI Python Package Index is also possible by executing

Statistik?

- Noch bisschen üben? – jupyter-tutorials
- Skript: Datenauswertung in den Grundlagenpraktika
- Skript: Funktionsanpassung mit der chi2-Methode



Name	Last commit	Last update
Fehlerrechnung.ipynb	initial commit	3 weeks ago
IntroStatistik.ipynb	initial commit	3 weeks ago
JupyterCheatsheet.ipynb	initial commit	3 weeks ago
Jupyter_Logo.svg	initial commit	3 weeks ago
PythonCheatsheet.ipynb	initial commit	3 weeks ago
PythonIntro.ipynb	initial commit	3 weeks ago
README.md	adjusted links	3 weeks ago
advancedFitting.ipynb	initial commit	3 weeks ago
jupyterTutorial.html	fixed broken links: ekp -> etp	1 week ago
jupyterTutorials.ipynb	initial commit	3 weeks ago
kafe2Tutorial.ipynb	initial commit	3 weeks ago
matplotlibTutorial.ipynb	initial commit	3 weeks ago
negLogLFits.ipynb	initial commit	3 weeks ago
pandasBasics.ipynb	initial commit	3 weeks ago

Linksammlung

- https://labs.physik.kit.edu/prakt-klass-physik.php

Erwartungshorizont

- Keine Messung ohne Unsicherheit!
- Wie bestimme ich Unsicherheiten?
- Wahl des Modells
- Computergestützte Fehlerfortpflanzung
- Schätzung von Parametern/Unsicherheiten
- Pull $\rightarrow$ verstehen und interpretieren
- χ^2-Test $\rightarrow$ verstehen und interpretieren

Mit diesem Rüstzeug sollten Sie fit sein, um den Vorversuch Datenverarbeitung am Beispiel des Pendels bestreiten zu können.
Nehmen Sie sich die Zeit den Vorversuch gut zu verstehen.

Nächste Termine

Datum	Zeit	Ort	Veranstaltung/Ereignis
Mo 23.10.	13:00 -- 14:30	Gaede-HS	Studenten Vorbesprechung
Do 26.10.	17:30 -- 19:00	Gaede-HS	Einführungsvorlesung Datenverarbeitung
Mo 30.10.	13:30 -- 19:00	P1	Erster Praktikumstag Mo-Gruppe **Gemeinsame Besprechung Vorversuch**
Mo 30.10.	**19:00 – 20:00**	**30.23 SR 9-1**	**Fragestunde zur Datenverarbeitung**
Do 02.11.	13:30 -- 19:00	P1	Erster Praktikumstag Do-Gruppe **Gemeinsame Besprechung Vorversuch**
Mo 06.11.	13:30 -- 19:00	P1	Zweiter Praktikumstag Mo-Gruppe **Fälligkeit Protokoll des Vorversuchs**
Do 09.11.	13:30 -- 19:00	P1	Zweiter Praktikumstag Do-Gruppe **Fälligkeit Protokoll des Vorversuchs**

Danach läuft das P1 nach Plan, wie auf den Webseiten zum P1 angegeben.

Backup

Erkenntnisgewinn seit Gallileis Zeiten

- Seit der Frühmoderne geleitet von mathematischen Modellen

Konstruktion des Modells

Erklärungskraft

Experiment:

Eintreten eines realen Ereignisses A

Modell:

Modellraum, Modellereignis A'

Überprüfung

Vorhersage

Eintreten des realen Ereignisses B

Mathematisches Resultat B'

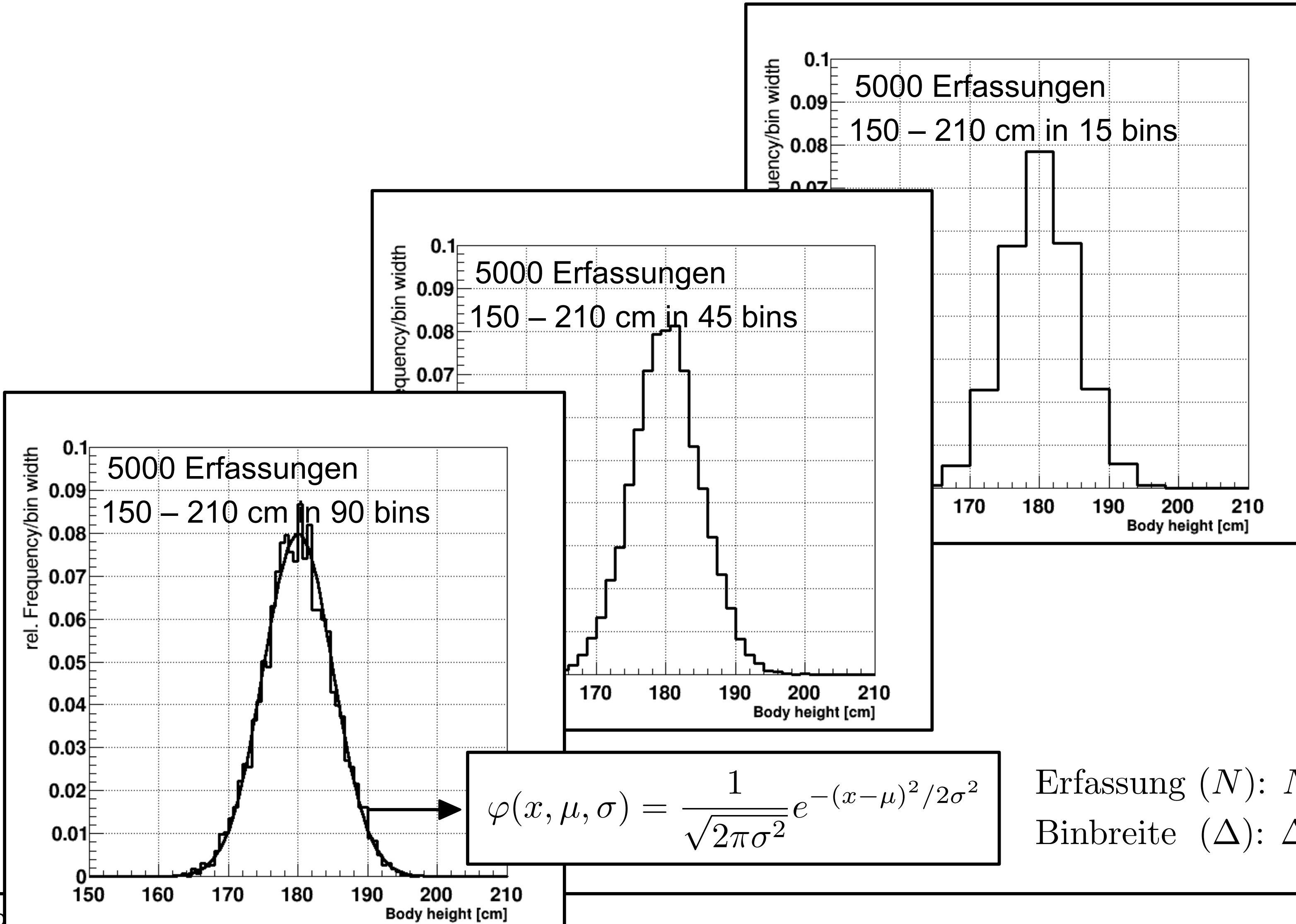
Übertragung in die Realität

Vorhersagekraft

- Einfache Erklärung & Vorhersage reproduzierbarer Naturereignisse

Histogramm → Wahrscheinlichkeitsdichte

- **Darstellung** durch Histogramm, dividiert durch Stichprobenlänge (=5000) und Binbreite.



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

Erfassung (N): $N \to \infty$

Binbreite (Δ): $\Delta \to 0$

Mittelwert der Stichprobe

Die Größe

$$\overline{x} = \frac{1}{n} \sum_{i \leq n} x_i$$

heißt Mittelwert der Stichprode. Ihr Erwartungswert ($E[\overline{x}]$) und ihre Varianz ($\mathrm{var}[\overline{x}]$) sind:

$$E[\overline{x}] = E\left[\frac{1}{n}\sum_{i\leq n} x_i\right] = \frac{1}{n}\sum_{i\leq n}\underbrace{E[x_i]}_{\equiv \mu} = \frac{1}{n}\sum_{i\leq n}\mu = \mu$$

$$\mathrm{var}[\overline{x}] = E[\overline{x}^2] - E[\overline{x}]^2 = E\left[\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)\left(\frac{1}{n}\sum_{j=1}^{n} x_j\right)\right] - \mu^2 = \frac{1}{n^2}\sum_{i,j=1}^{n} E[x_i x_j] - \mu^2$$

$$= \frac{1}{n^2}\left[\underbrace{(n^2-n)\mu^2}_{} + \underbrace{n(\mu^2+\sigma^2)}_{}\right] - \mu^2 = \sigma^2/n\,.$$

Dabei sind μ und σ^2 der Erwartungswert und die Varianz der Einzelmessung und n die Länge der Stichprobe.

$n(n-1)$ "off-diagonale" Elemente
$(i \neq j)$ mit:
$E\left[x_i x_j\right] = E\left[x_i\right] E\left[x_j\right] = \mu^2$

n "diagonale" Elemente
$(i = j)$ mit:
$E\left[x_i x_i\right] = \mu^2 + \sigma^2$

Kovarianz

Analog zur Varianz definiert man die Kovarianz

$$\mathrm{cov}[x,y] = E\left[(x - E[x])\,(y - E[y])\right] = E[x\,y] - E[x]E[y]$$

zur Beschreibung der Beziehung zweier Zufallsvariablen x und y zueinander. Die Matrix

$$V_{xy} = \begin{pmatrix} \sigma_x^2 & \mathrm{cov}[x,y] \\ \mathrm{cov}[x,y] & \sigma_y^2 \end{pmatrix}$$

bezeichnent man als Kovarianzmatrix. Die Größe

$$\rho_{xy} \equiv \frac{\mathrm{cov}[x,y]}{\sigma_x \sigma_y}$$

als (linearen) Korrelationskoeffizienten.

- V_{xy} is symmetrisch, d.h. es gibt immer eine Hauptachsentransformation und reelle Eigenwerte
- $\rho_{xy} \in [-1,+1]$. Für unabhängige Zufallsvariablen gilt:

$$E[\,x\,y\,] = E[\,x\,]E[\,y\,] \quad \text{d.h.} \quad \rho_{xy} = 0$$

Priv.-Doz. Dr. Roger Wolf
http://ekpwww.physik.uni-karlsruhe.de/~rwolf/

Korrelationskoeffizient der Stichprobe

Die Größe

$$r = \frac{\hat{V}_{xy}}{s_x s_y} = \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{\sum (x_j - \overline{x})^2 \sum (y_k - \overline{y})^2} = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\sqrt{(\overline{x^2} - \overline{x}^2)(\overline{y^2} - \overline{y}^2)}}$$

ist eine Schätzfunktion für den Korrelationskoeffizienten zweier einzelner Zufallsvariablen x und y. Sie hat den Erwartungswert und die Varianz:

$$E[r] = \rho - \frac{\rho(1-\rho^2)}{2n} + \mathcal{O}(1/n^2)$$

$$\mathrm{var}[r] = \frac{1}{n}(1-\rho^2)^2 + \mathcal{O}(1/n^2)$$

Priv.-Doz. Dr. Roger Wolf
http://ekpwww.physik.uni-karlsruhe.de/~rwolf/