Tagebuch 26.2.

Heute Kapitel 6: "Umgang mit Datensätzen" begonnen.
Besprochen wurde Kap. 6.1 und Kap. 6.2 sowie die erste Computersimulation
Verstanden zu **merkende Formeln:**
Mittelwert, Streung der Daten um das Datenmittel
Schätzung der Datenstreuung um den wahren Wert
Schätzung der Streuung der Mittelwerte um den wahren Wert!

Nachfolgend ein eigener Datensatz zur Pendelschwingungsdauer. In der ersten Spalte die wie angegeben vermessene Pendellänge (in cm) und rechts die Dauer von 20 Schwingungen (in s).

$$D = \begin{array}{cc} 10 & 14.5 \\ 020.5 & 20.25 \\ 34 & 24.5 \\ 47 & 28.75 \\ 62.5 & 32.5 \\ 69 & 34.25 \\ 73 & 35.75 \\ 85 & 38.75 \\ 113 & 43 \\ 155.5 & 50 \end{array}$$

In diesem Datensatz wurden die letzten drei Längen etwas ungenau und unachtsam vermessen.

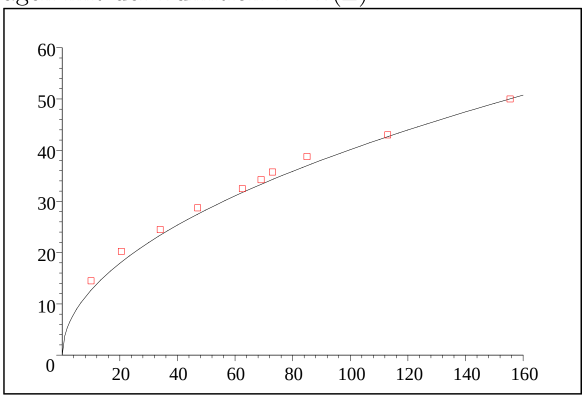
Von Ihrer Seite wurde - vollkommen zu recht - vorgeschlagen, nach der Pendelläne aufzulösen und diese soden über T und g=9.81cm/s^2 effektiv zu besimmen. Das gibt mit $\boxed{\mathrm{L}=\left(\frac{T}{2\pi\cdot 20}\right)^2 981}$ folgende Liste effektiver Längen, von der wir sogleich die gemessenen Längen subtrahieren:

$$\begin{array}{c} 13.061 \\ 25.474 \\ 37.289 \\ 51.349 \\ 65.617 \\ 72.873 \\ 79.396 \\ 93.281 \\ 114.86 \\ 155.31 \end{array} - \begin{array}{c} 10 \\ 020.5 \\ 34 \\ 47 \\ 62.5 \\ 69 \\ 73 \\ 85 \\ 113 \\ 155.5 \end{array} =: \begin{array}{c} 3.061 \\ 4.974 \\ 3.289 \\ 4.349 \\ 3.117 \\ 3.873 \\ 6.396 \\ 8.281 \\ 1.86 \\ -0.19 \end{array}$$

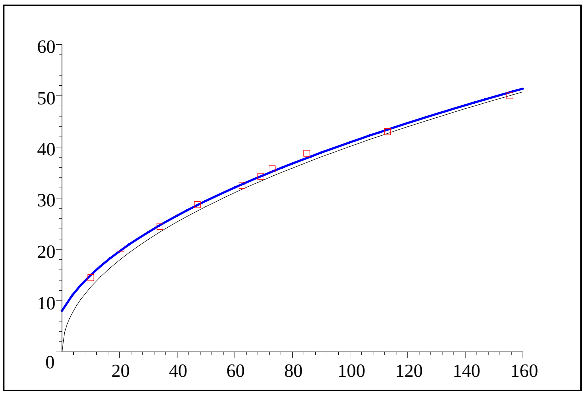
Und das heißt, die vermessenen Längen sind alle etwa 3-4cm zu kurz. Das zugehörige mathematische Pendel hat eine etwa größere Längen. Mittelwert 3.9

Verschiedene graphische Auftragungen

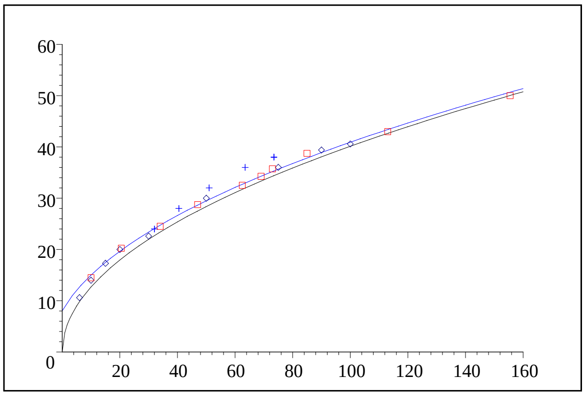
Die daten einfach aufgetragen mit der Funktion T=T(L)



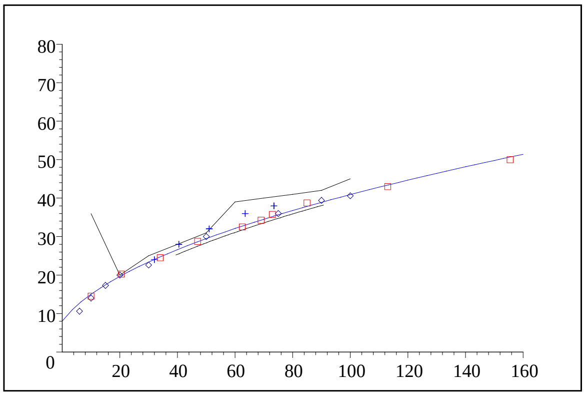
Und jetzt mit mit T=$20 \cdot 2\pi\sqrt{\frac{L-4}{g}}$. Man sieht die verbesserte Übereinstimmung



Zwei Datensätze wurden mir von Ihrer Seite abgeliefert. diese sind in der nachfolgenden Figur mit eingetragen:



Man sieht auch hier: Die Werte sind systematishh zu hoch. Oder auch: Das mathematische Ersatzpendel mus eine etwa 3-4cm größere Länge haben. Und das ist deutlich größer als der Abstand des Aufhängepunktes zum Schwerpunkt, der etwa 2cm beträgt.



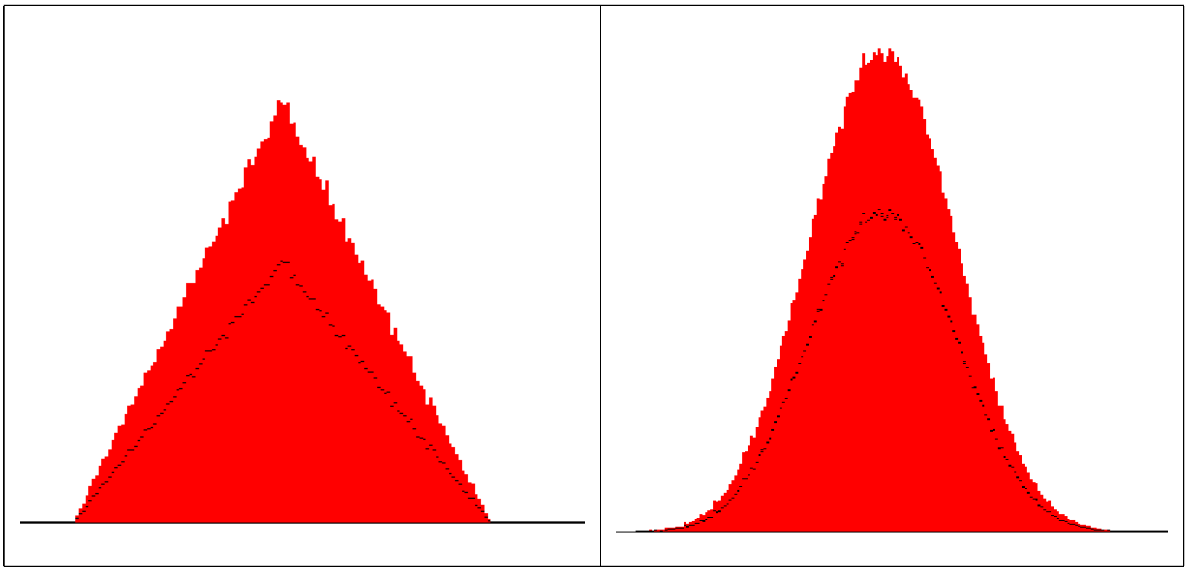
27.2.

Verstanden zu **merkende Formeln aus Kap. 6:**
Mittelwert, Streung der Daten um das Datenmittel
Schätzung der Datenstreuung um den wahren Wert
Schätzung der Streuung der Mittelwerte um den wahren Wert!

Es werden zwei gleichverteilte Zufallszahlen zwischen 0 und 1 gezogen und deren Summe gebildet. Dieser Vorgang wird N-mal wiederholt. Das ergibt einen Datensatz vom Auszählungstyp.

a) Welche Werte sind möglich? (Alle zwischen 0 und 1)
b) Was für eine Verteilung der Werte ist zu erwarten? ("Dreiecksverteilung")
b1) Bei einer mittigen Einteilung des Werteintervalles in 2 Teile
b2) In sehr viele Teile
b3) Wie sollte die Verteilung der relativen Häufigkeiten für sehr große N (N$\rightarrow \infty$) aussehen?
c) Was ändert sich, wenn man die Summe von drei derartigen Zufallszahlen bildet?

▼ Hierzu zwei Bilder, einmal für die Summe aus zwei, einmal aus drei Zufallszahlen:



In beiden Fällen erolgt eine Auszählung in 200 Häuser. Schwarz noch die relativen Häufigkeiten, die sich mit wachsendem N nur wenig ändern.

Wir gehen aus von der Linsenformel und nehmen an, dass die Brennweite f der Linse nur ungenau bekannt sei: $f_0 \pm \Delta f$. Dagegen sei g genau bekannt. Welche Ungenauigkeit bewirkt das für die Bildweite? Wie kann man diese Ungenauigkeit näherungsweise bestimmen?

f ist veränderlich, also unabh. Variable. Abhängig ist b. Also die Linsenformel *nach* b umstellen zu b=$b_g(f)$. Die von der Ungenauigkeit von f bewirkte ungenausigkeit in b ist gegeben durch

$$b_g(f_0 \pm \Delta f) - b_g(f_0). \approx \pm \frac{db_g}{df}(f_0)\Delta f.$$

wobei wir näherungsweise die mittlere Änderungsrate durch die momentane ersetzen können. Rechnerisch ist

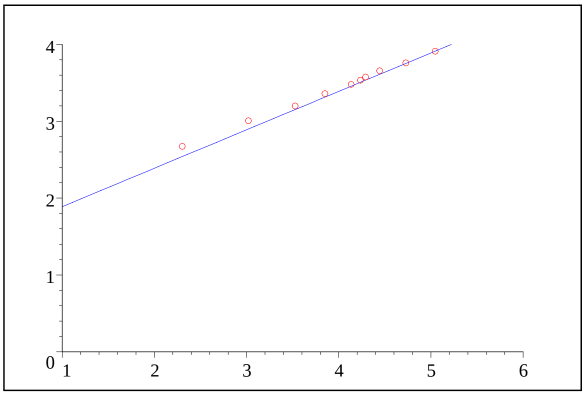
$$b = b(f) = \frac{gf}{g-f} \qquad \frac{db}{df}(f) = \frac{g(g-f) - gf(-1)}{(g-f)^2} = \frac{g^2}{(g-f)^2}.$$

Wir gehen aus von der Formel für die Schwingungsdauer des Pendels. Bilden Sie von beiden Seiten der Gleichung den Logarithmus.

$$T = 2\pi g^{-\frac{1}{2}} L^{\frac{1}{2}} \qquad \underbrace{\log(\mathrm{T})}_{y}=\log(2\pi g^{-\frac{1}{2}}) + \frac{1}{2}\underbrace{\log(L)}_{x}$$

Was erwarten Sie daher, wenn man einen Datensatz (Pendellänge, zugeh. Schwingungsdauer) doppelt logarithmisch aufträgt? Was für ein Graph entsteht? Der einer Gerade mit Steigung $\frac{1}{2}$

Die folgende figur zeigt das Resultat für den am Vortag gegebenen Datensatz. Mit eingetragen ist die vorhergesagte Gerade $\log\mathrm{T}=\mathrm{b}+\frac{1}{2}\log L$ mit $\mathrm{b}=\log\frac{20\cdot 2\pi}{\sqrt{981}}$. (Die Daten sind in cm und Sekunden angegben.)



Eine Aufgabe. zum relativem Fehler.

□ Wir ersetzen sin(x) durch x.

a) Was für ein absoluter Fehler wird bei dieser Ersetzung gemacht ($\boxed{\mathrm{AF}=\sin(x)\text{-}x}$ ...)

b) Bestimmen Sie die beiden relativen Fehler dieser Ersetzung: $(\mathrm{RF}_W = \frac{\sin(x)-x}{\sin(x)}$.., $\mathrm{RF}_S = \frac{\sin(x)-x}{x})$

Bemerkung: Gemeint ist in der Regel die erste Größe. Meist hat man aber Werte zur Bestimmung der zweiten. Dann nimmt man diese als Schätzung für die erste. Der Unterschied ist meist nicht groß und für die üblichen Fragestellungen unwichtig. Index W für "wahr" und S für "Schätzung".).

c) Welcher Fehler wird gemacht, wenn man RF_W durch RF_S ersetzt?

$$\boxed{\mathrm{AFF}=\frac{\sin(x)-x}{\sin(x)} - \frac{\sin(x)-x}{x} = -\frac{-2(\sin x)x+x^2+\sin^2 x}{(\sin x)x}}..$$

d) Einsetzen einiger Zahlwerte. Wähle x=0.3 und x=0.8

$$\begin{array}{cc} \sin .3 - .3 & \sin .8 - .8 \\ \frac{\sin(.3)-.3}{\sin(.3)} & \frac{\sin(.8)-.8}{\sin(.8)} \\ \frac{\sin(.3)-.3}{.3} & \frac{\sin(.8)-.8}{.8} \\ \frac{\sin(.3)-.3}{\sin(.3)} - \frac{\sin(.3)-.3}{.3} & \frac{\sin(.8)-.8}{\sin(.8)} - \frac{\sin(.8)-.8}{.8} \end{array} = \begin{array}{cc} -4.4798 \times 10^{-3} & -8.2644 \times 10^{-2} \\ -1.5159 \times 10^{-2} & -0.11521 \\ -1.4933 \times 10^{-2} & -0.1033 \\ -2.2636 \times 10^{-4} & -1.1901 \times 10^{-2} \end{array}$$

Die logarithmische Auftragung des Einzelfehlers (als Funktion der Schrittweite Δx wurde diskutiert und in einer animation gezeigt!

Einzelfehler $\boxed{\text{EF}=\text{c}\Delta x^2}$

$\boxed{\log(\text{EF})}=\log(\text{c})+2\boxed{\log(\Delta x)}$

In der ersten Aufgabe von Freitag lag ein Schreibfehler vor, den man auf zwei Weisen interpretieren konnte. Ein klammerzeichen fehlte. Wir bestimmen die Ableitung für beide Interpretationen.

Die erste richtige Interpretation:

$$\vec{r}_2(t) = \begin{pmatrix} \cos(\varphi(t)) \cdot \sin(\theta(t)) \\ \sin(\varphi(t)) \cdot \sin(\theta(t)) \\ \cos(\theta(t)) \end{pmatrix}$$

Rechnung, ohne Zwischenrechnung. Zuvor Strategieüberlegung über die form des Ergebnisses und den Weg dahin: :

$$\boxed{\vec{v}_2(t) = \dot{\theta}(t) \begin{pmatrix} \cos\theta(t)\cos\varphi(t) \\ \cos\theta(t)\sin\varphi(t) \\ -\sin\theta(t) \end{pmatrix} + \dot{\varphi}(t)\sin\theta(t) \begin{pmatrix} -\sin\varphi(t) \\ \cos\varphi(t) \\ 0 \end{pmatrix}}$$

Und jetzt die ander nicht gemeinte Interpretation:

$$\vec{r}_2(t) = \begin{pmatrix} \sin(\theta(t) \cdot \cos\varphi(t)) \\ \sin(\theta(t) \cdot \sin\varphi(t)) \\ \cos\theta(t) \end{pmatrix}$$

Strategieüberlegung: Erst innere Ableizung $\times$ dann äußere:

$$\vec{v}(t) = \begin{pmatrix} \cos(\theta(t) \cdot \cos\varphi(t)) \cdot \left[\dot{\theta}(t)\cos\varphi(t) + (-\dot{\varphi}(t))\sin\varphi(t)\right] \\ \cos(\theta(t) \cdot \sin\varphi(t)) \cdot \left[\dot{\theta}(t)\sin\varphi(t) + (\dot{\varphi}(t))\cos\varphi(t)\right] \\ -\dot{\theta}(t)\sin\theta(t) \end{pmatrix}$$

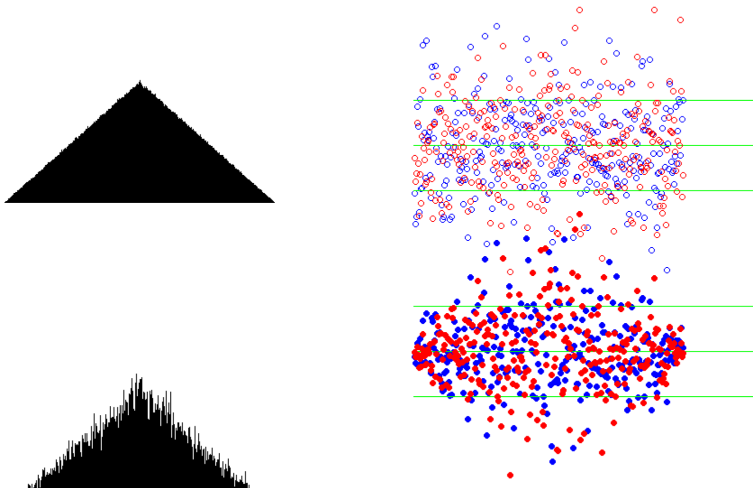
Stoff:

Das wichtige Kapitel 6.6 durchgesprochen! Wie genau schätzt eine absolute Häufigkeit einen wahren Wert?

Sei N_i der ausgezählte Wert und $\text{p}_i N$ der wahre oder der vorhergesagte Wert. Bilde dazu die Koordinatengröße $\frac{N_i - p_i N}{\sqrt{N_i}}$ bzw deren Quadrat. Bedeutung dieser Größe?

Dazu eine Computeranimation gezeigt:

Die oberen Punkte der Animation bestimmten diese Werte. Sie liegen immer im und in der Nähe des Intervalles -2<w<2. Die animation zeigt, dass das für ganz unterschiedliche Werte von N un N_i der Fall ist. In der unteren Verteilung ist N=20000 und in der oberen ist N=2000000. Unten rechts sind die absoluten Fehler aufgetragen, wobei die zur oberen Verteilung gehörigen um einen Faktor $10=\sqrt{100}$ herabskaliert sind Oben sind die Koordinatengrößen $\frac{N_i - p_i N}{\sqrt{N_i}}$ für beide Fälle im Bereich -1<w<1 aufgetragen (grüne Linien!)



28.2.

Zunächst: Einige Warmdenkübungen gerechnet. Erneut das Problem: Die zusammengefassten Ergebnisse des Vortages sollten verfügbar sein. Klappte nicht. Probleme bei der vierten aufgaben

Dann das Thema lineare Regression nach Skript. Wie macht man sich einen solchen Text zugänglich, insbesondere die Resultate . Geübt: Formel oder Rechnung anschauen, abdecken und selbst rekonstruieren.

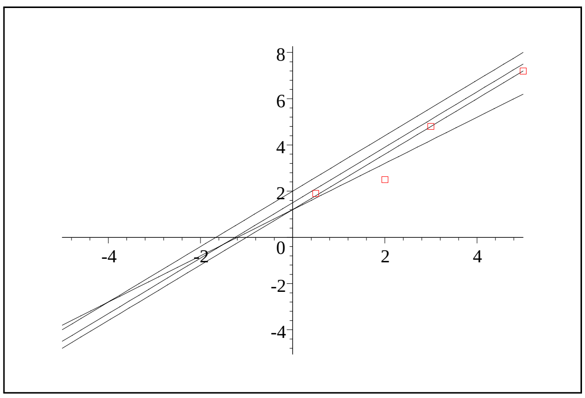
Nach der Pause: Nochmals die einleitende Zusammenfassun des Kurses durchgegangen, was ist zu merken, wie geht man an die Probleme heran. Sah etwas besser aus, aber nach einiger Zeit machten sich Ermüdungserscheinungen bemerkbar!

☐ 1)Gegeben ein Datensatz a von Zahlpaaren:

1	2	3	4
(0.5,1.6)	(2,3.2)	(3,4.0)	(5, 6.3)

a) Skizzieren Sie den Datensatz als "Punkteschwarm" in der Ebene,

b) Zeichnen Sie "per Augenmaß" eine passende Gerade hinein und schätzen Sie die Werte für m und b. (Lineal, kariertes Papier, Schnell! 2-3 Minuten)



c) Bestimmen Sie rechnerisch die Lage des Schwerpunktes und zeichnen Sie ihn mit ein.

☐ 2) Sie messen die jeweilige Länge von N=475. Werkstücken aus. Sie finden, dass davon n=25 einen zu großen Wert haben. Erlaubt ist eine relative Häufigkeit $h_A = 0.08$...Ist der von Ihnen gefundene Wert akzeptabel oder nicht? E**rst** allgemeine Formel, dann die speziellen Zahlwerte!

▼Man bnötigte an bestandenen Formeln bzw. Ausdrücken: $N \pm \sqrt{N}$ und $\Delta N = h_A \cdot N$.▲

☐ 3) Sie messen die jeweilige Länge von N=475. Werkstücken aus. Sie finden, dass davon n=25 einen zu großen Wert haben. Was erwarten Sie nach der Fluktuationsregel, für eine Messung weiterer 475 Werkstücke? Was für eine Messung von 1000 Werkstücken?

☐ 4) Sie haben eine vierfache Messung einer gesuchten Größe x_W. Die Resultate sind

1	2	3	4
5.31	5.27	5.33	5.25

Was können Sie über den Wert x_w aussagen, wenn Sie annehmen, dass kein systematischer Fehler vorliegt?

▼ Mittelwert / Für die Ungenauigkeit: $\sigma^S_{MW} = \frac{\sigma^S_{Daten,wahr}}{\sqrt{N}}$ Antwort: $\boxed{5.29 \pm 0.018}$ ▲

.

Herleitung der Formeln für die lineare Regression: Strategie und Notizen:

$$\begin{aligned}\frac{\partial A}{\partial m}(m,b) &= -2\sum_{i=1}^{N} x_i(y_i - mx_i - b) \stackrel{!}{=} 0 \\ \frac{\partial A}{\partial b}(m,b) &= -2\sum_{i=1}^{N} (y_i - mx_i - b) \stackrel{!}{=} 0\end{aligned}$$

$$\sum x_i = N\vec{x} \qquad \sum x_i y_i = N(\vec{x}\vec{y}) \qquad \sum x_i^2 = N\vec{x}^2$$

$$\begin{aligned}\sum x_i y_i &= m\sum x_i^2 + b\sum x_i \\ \sum y_i &= m\sum x_i. + bN...\end{aligned}$$

$$\begin{aligned}(\vec{x}\vec{y}) &= m\vec{x}^2 + b\bar{x} \qquad (1) \\ \bar{y} &= m\bar{x} + b \qquad (\text{-}\bar{x}) \\ m &= \frac{(\vec{x}\vec{y}) - \bar{x}\bar{y}}{\vec{x}^2 - \bar{x}^2}\end{aligned}$$

☐ Wie groß ist der relative Fehler, den man macht, wenn man anstatt der Schätzung der Streuung um den wahren Wert die übliche Datensatzstreuung nimmt?

▼ Es muss N>1 gelten!

$$\begin{aligned}RF &= \frac{\sigma^S_{Daten,wahr} - \sigma}{\sigma^S_{Daten,wahr}} = \frac{\sqrt{\frac{1}{N-1}\sum(a_i - \bar{a})^2} - \sqrt{\frac{1}{N}\sum(a_i - \bar{a})^2}}{\sqrt{\frac{1}{N-1}\sum(a_i - \bar{a})^2}} \\ &= \frac{\sqrt{\frac{1}{N-1}} - \sqrt{\frac{1}{N}}}{\sqrt{\frac{1}{N-1}}} = \frac{\sqrt{N} - \sqrt{N-1}}{\sqrt{N}} = 1 - \sqrt{\frac{N-1}{N}} \\ &= 1 - \sqrt{1 - \frac{1}{N}}\end{aligned}$$

1.3.2007
Die letzten heute besprochenen Aufgaben:

Und wieder zeigt sich, dassviele die vorangegangenen Ergebnisse vergessen hatten. Was ist eine relative Größe? Was eine Auszählung. Wie erhält man numerisch die lösung einer Differentialgleichung? Welcher Unterschied besteht zwischen Datenstreuung und Streuung der Mittelwerte. usw.

Die heute besprochenen Aufgaben in nicht ausgearbeiteter Form!

Konzept - Endformulierung

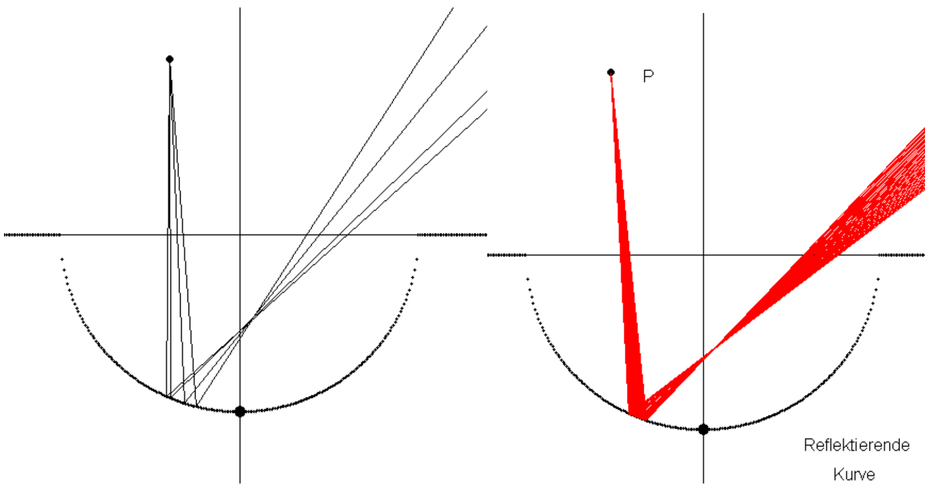
□ Nachfolgend eine Auszählung eines Datensatzes der ersten Ziffer der Einwohnerzahl von Ländern der Welt.

1	2	3	4	5	6	7	8	9
53	27	24	18	18	6	14	13	7
0.29	0.15							
7. 3								

a) Bestimmen sie die zugehörigen relativen Häufigkeiten N=180
b) Zeichnen Sie ein Histogramm der Verteilung der absoluten Zahlen
c) Bestimmen Sie die Größe der zugehörigen 1-σ−Fluktuationen und zeichnen Sie diese in die Skizze ein

---

□ Die nachfolgende Skizze erfasst einen Sachverhalt aus der Optik.
a) Was wird dargestellt?
b) Welches zu lösende mathematische Problem entsteht, wenn der Ort des Punktes P, das davon ausgehende schmale Lichtbündel ebenso wie die Form der reflektierenden Kurve vorgegeben ist?
c) Formulieren Sie eine Strategie, wie dieses Problem mit Hilfe uns bekannter mathematischer Methoden zu losen ist.
d) (Angenommen die Kurve ist ein Kreis mit Radius R und das Bündel liegt auf der opt. Achse). Dann kennen Sie das Resultat. Nämlich?)



☐ Flugparabel: Ein Massenpunkt bewegt sich auf einer Flugparabel. Es gilt $\vec{r}^K(-3) = \begin{pmatrix} 0 \\ 0 \\ h \end{pmatrix}$ und $\vec{v}^K(-3) = \begin{pmatrix} 0 \\ 2 \\ 3 \end{pmatrix}$ und $\vec{g}^K = \begin{pmatrix} 0 \\ 0 \\ -g \end{pmatrix}$.

Bestimmen Sie den Ort des Scheitelpunktes (in Form einer Formel).
Zunächst a) Wie sieht die Vorgehensstrategie aus?

$$\vec{r}(t) = \vec{r}_1 + \vec{v}_1 T + \tfrac{1}{2}\vec{g}T^2 \qquad \text{T=t-t}_1 \quad \text{hier } \boxed{\boxed{\text{T=t+3}}}$$

$$\vec{r}^K(t) = \begin{pmatrix} 0 \\ 0 \\ h \end{pmatrix} + \begin{pmatrix} 0 \\ 2 \\ 3 \end{pmatrix} T + \frac{1}{2}\begin{pmatrix} 0 \\ 0 \\ -g \end{pmatrix} T^2 = \begin{pmatrix} 0 \\ 2T(t) \\ h + 3T(t) - \frac{1}{2}gT^2(t) \end{pmatrix}$$

$$\vec{v}^K(t) = \begin{pmatrix} 0 \\ 2 \\ \boxed{\text{3-gT}} \end{pmatrix}$$

Scheitel: $\text{T}_S = \frac{3}{g} \quad \text{t}_s = \frac{3}{g} - 3$

$$\vec{r}^K(\tfrac{3}{g} - 3) = \begin{pmatrix} 0 \\ \frac{6}{g} \\ h + \frac{9}{2g} \end{pmatrix} \qquad -\frac{1}{2}g\frac{9}{g^2}$$

☐ Numerisches Lösen der Differentialgleichungen
Gegeben die Differentialgleichung $\boxed{y'(x) = y^2(x) + x^2}$ mit dem Anfangswert y(1)=1.
Bestimmen Sie den Einzelschrittfehler, den Sie im ersten Schritt für $\Delta x = 0.2, 0.1, 0.05$ und 0.025 machen und kommentiern Sie das Resultat. Wieso sollte man wohl log(Einzelfehler) gegen log(Δx) auftragen.

- Wieso ist diese Aufgabe so nicht lösbar???
- Formulieren Sie das Argument, das zeigte, dass der Gesamtfehler proportional zu Δx ist.
- Wie sieht die Grundformel für die numerische ´Methode aus und das dasraus folgende REchenschema? Formel y(x+Δx) = ... Und das Schema für einen Schritt:

x	1	$1 + \Delta x$
y	1	1.4
$\frac{dy}{dx}(x_0)$	2	
$\frac{dy}{dx}(x_0)\Delta x$	0.4	

x	1	$1 + \Delta x$
y	1	1.2
$\frac{dy}{dx}(x_0)$	2	
$\frac{dy}{dx}(x_0)\Delta x$	0.2	

Rechenbeispiel zur Bestimmung einer Regressionsgeraden.
Der Datensatz: .... Die Formeln für m und b werden aus dem Skriptum genommen. Dann sind eine Reihe von Hilfsgrößen zu berechnen:

$$
\begin{aligned}
\bar{x} &= \frac{1}{4}(.5+1+2+3) = 1.625 \\
\bar{y} &= \frac{1}{4}(1.1+1.9+4.3+5.5) = 3.2 \\
\vec{x}^2 &= \frac{1}{4}(\frac{1}{4}+1+4+9) = 3.5625 \\
\vec{y}^2 &= \frac{1}{4}(1.1^2+1.9^2+4.3^2+5.5^2) = 13.39 \\
\vec{x}\vec{y} &= \frac{1}{4}\left(.5\cdot 1.1+1\cdot 1.9+2\cdot 4.3+3\cdot 5.5\right) = 6.8875
\end{aligned}
$$

$$: \ x^2 =$$

$$\tfrac{1}{4}(1.1^2+1.9^2+4.3^2+5.5^2) = 13.39$$

$$\tfrac{1}{4}\left(.5\cdot 1.1+1\cdot 1.9+2\cdot 4.3+3\cdot 5.5\right) =$$

$$\text{m}=\tfrac{6.8875-3.2\cdot 1.625}{3.5625-1.625^2} = 1.8305$$
