



Introduction to probability and statistics

Piotr Nowak

piotr.nowak@agh.edu.pl

About the course

- Lectures = seminars
- During the course we will discuss:
 - Measurements and uncertainties,
 - Classical probability
 - Probability distribution (discrete and continuous)
- Two tests - to pass the course, it is advisable to get 50% of the total number of points from both tests

Introduction to probability and statistics

Lecture 1.

Uncertainty in measurements



MEASUREMENT

The result of a measurement is only an approximation or estimate of the value of the specific quantity subject to measurement, the measurand which can be classified as:

- ☐ simple, or
- ☐ complex

Example: Mathematical pendulum, l – the length, T – period are simple measurands; measured directly

$$T = 2\pi\sqrt{\frac{l}{g}}$$

Determination of gravitational acceleration : g -complex measurand

MEASUREMENT

$$g = \frac{4\pi^2}{T^2} l$$

Let us assume:

- 1) $l = 2 \text{ m}$, $T = 2,84 \text{ s}$
- 2) $l = 2 \text{ m}$, $T = 0,227 \text{ s}$
- 3) $l = 2 \text{ m}$, $T = 113,5 \text{ s}$

$$g = 9.78933 \frac{\text{m}}{\text{s}^2}$$

$$g = 1532.279594 \frac{\text{m}}{\text{s}^2}$$

$$g = 0.006129 \frac{\text{m}}{\text{s}^2}$$

Significant digits

Please write following numbers with the given accuracy:

- To the first two significant digits:

$$134.232 = 130$$

$$0.34242 = 0.34$$

$$0.0002425 = 0.00024$$

- To the first three significant digits:

$$1\,231.2032 = 1\,230$$

$$5\,223\,113.3 = 5\,220\,000$$

$$2.000121 = 2.00$$

Uncertainty

Practically, we do not know real values and estimate **uncertainties**, due to dispersion of results, from the laws of **statistics**.

Uncertainty is

- a **parameter** related to the result of measurements,
- characterized by **dispersion**
- assigned to the measurand in a **justified way**.

Measures of uncertainty

Uncertainty

systematic

- *related to the measuring instrument /device*
- *most commonly associated with the scale interval of the instrument*

random (statistic)

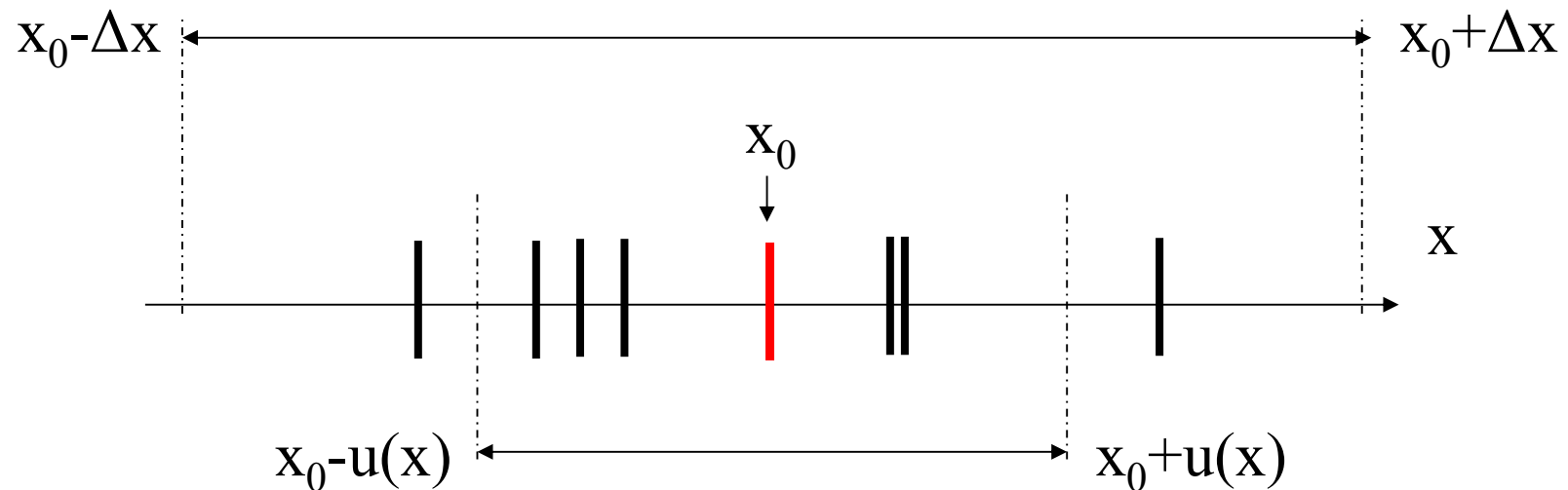
- *is associated with random scattering of data*

Measures of uncertainty

Uncertainty

maximum, Δx

standard, $u(x)$



Measures of uncertainty

In the simplest cases, the systematic uncertainty is the maximum uncertainty, and the random uncertainty is the standard uncertainty

Standard uncertainty

Distribution of random variable x_i , with a dispersion around the average \bar{x} is characterized by **standard deviation** defined as:

$$\sigma = \lim_{n \rightarrow \infty} \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$$

But ...

Standard uncertainty

...

Exact values of standard deviation are unknown [we cannot take an infinite number of measurements].

Standard uncertainty represents an estimate of **standard deviation**.

Standard uncertainty could be expressed as:

$$u(x) = \sigma_{AV} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n(n-1)}}$$

Example:

We have performed a series of measurements getting the following results x_1, x_2, \dots, x_n . In such a sample that can be considered as big some of the results are the same; n_k is a number of random experiments, in which the same result x_k has occurred.

x_k	n_k
5.12	1
5.17	1
5.22	2
5.24	4
5.27	7
5.33	10
5.36	14
5.46	16
5.52	13
5.61	8
5.64	6
5.68	4
5.73	3
5.79	1
Sum	90

Analysis of data

Arithmetic average

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

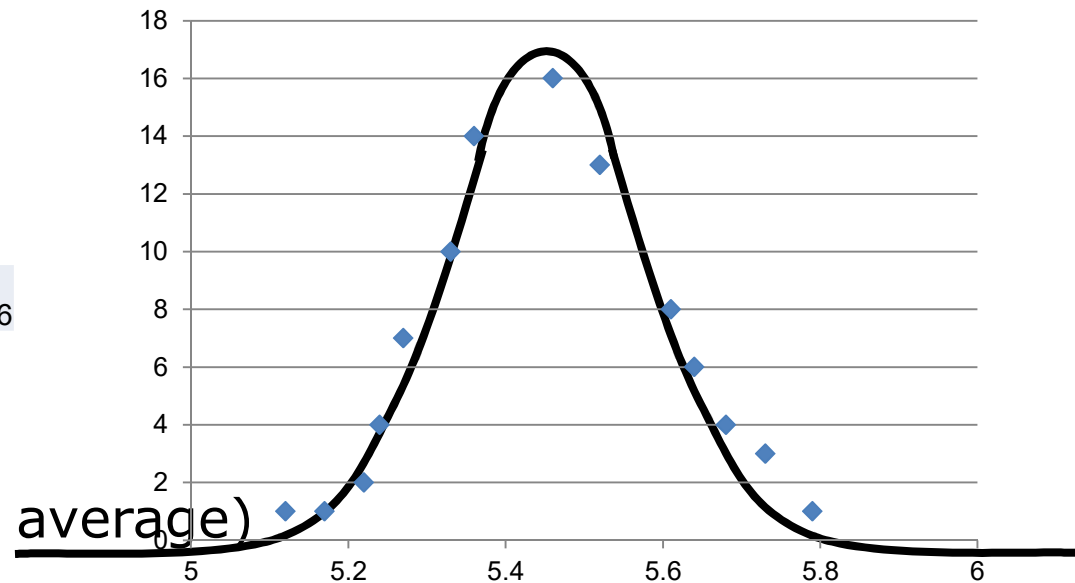
$\bar{X} = 5.449556$

Standard uncertainty

(standard deviation of the average)

$$u(x) = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n(n-1)}}$$

0.050083



Notation

If the maximum uncertainty is given:

$$\begin{array}{ccc}
 x=135.6 & \longrightarrow & x = (135.60 \pm 0.01) \\
 \Delta x = 0.01 & &
 \end{array}$$

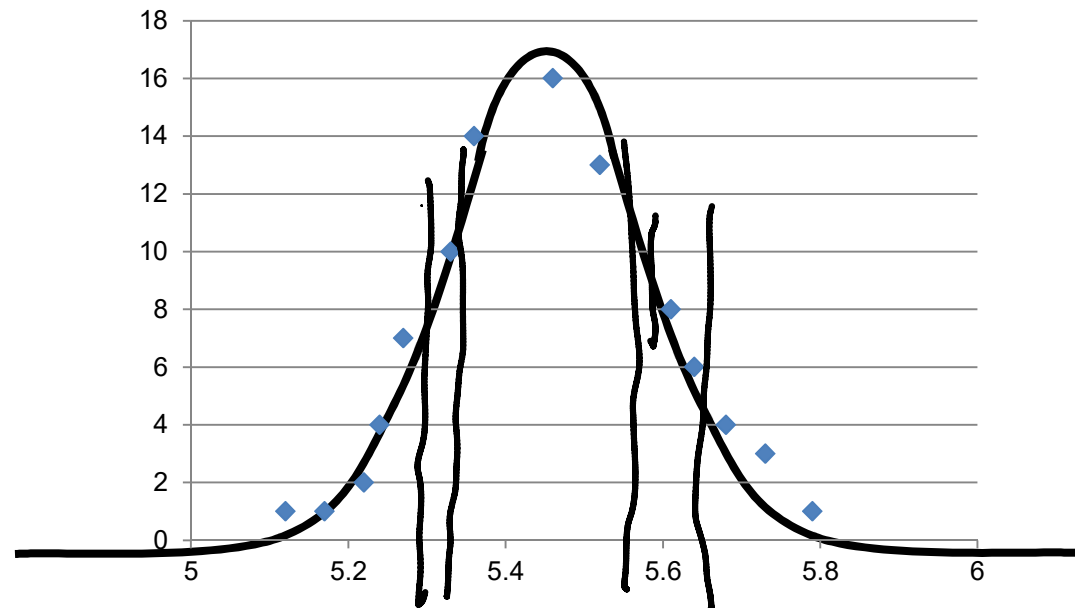
When we determine the standard uncertainty:

$$\begin{array}{ccc}
 x=135.2543232 & x=135.2543232 & \\
 u(x)=0.00142456 & u(x)=0.0015 & \searrow \\
 & & x=135.2543 \\
 & & u(x)=0.0015
 \end{array}$$

Analysis of data

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad \bar{x} = 5.449556$$

$$u(x) = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n(n-1)}} \quad 0.050083$$



$$u(x) = 0.050083 = 0.050 ; \quad x = 5.449556 = 5.450$$



$$x = 5.450$$

$$u(x) = 0.050$$

$$[\text{alternatively } x = (5.450 \pm 0.050)]$$

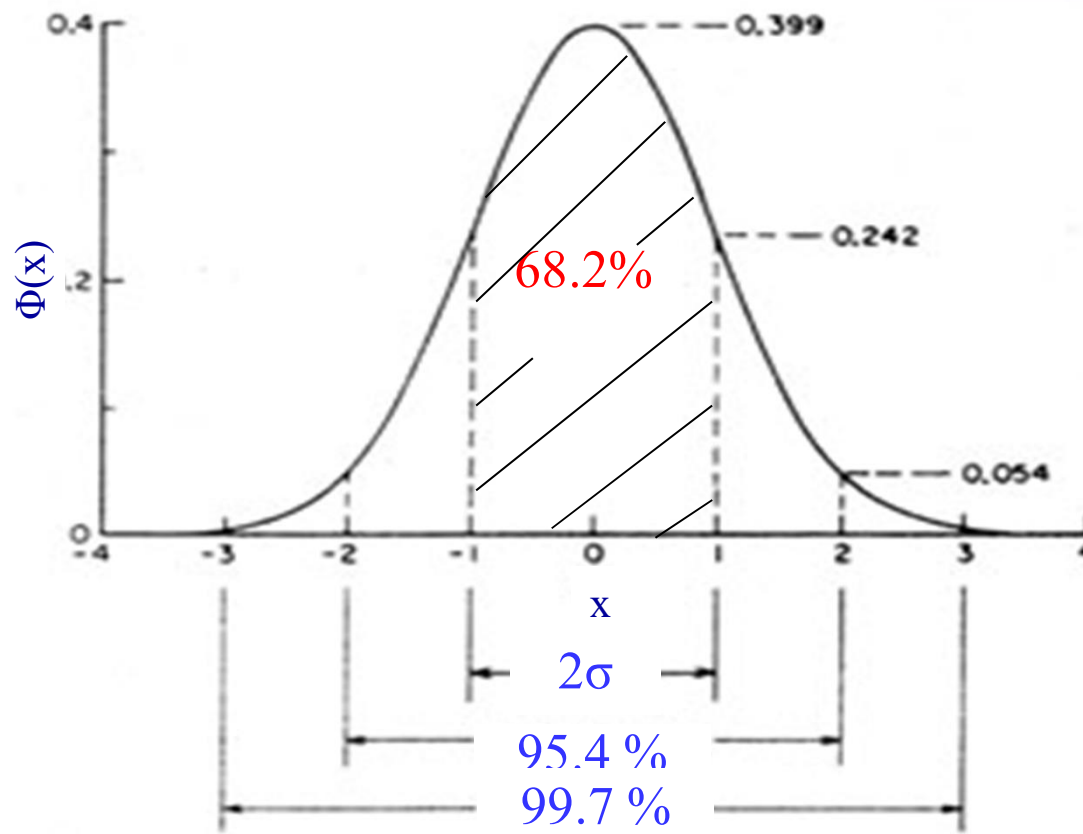
Gauss distribution function

Probability density function for the result x or its error Δx according to Gauss

$$\Phi(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left(-\frac{(x - x_0)^2}{2\sigma^2} \right)$$

x_0 is the most probable result and can be represented by the arithmetic average, σ is standard deviation, σ^2 is variance

Normal distribution



Within the interval $x_0 - \sigma < x < x_0 + \sigma$ we find 68.2 % (2/3),
 For $x_0 - 2\sigma < x < x_0 + 2\sigma$ - 95.4 %
 For $x_0 - 3\sigma < x < x_0 + 3\sigma$ - 99.7 %
 of all results

Maximum uncertainty vs. standard uncertainty

Maximum uncertainty - within this interval:

$$x_0 - \Delta x < x_i < x_0 + \Delta x$$

(almost) all the results x_i , will fall.

Standard uncertainty - within this interval:

$$x_0 - u(x) < x_i < x_0 + u(x)$$

about 68% the results x_i , will fall.

Maximum uncertainty vs. standard uncertainty

If we want to transform the standard uncertainty $u(x)$ into the maximum uncertainty $D(x)$:

$$\Delta x \approx 2 \cdot u(x)$$

(95% results are in $[x_0 - 2u(x) ; x_0 + 2u(x)]$)

$$u(x) = 0.050083 = 0.050 ; \quad x = 5.449556 = 5.450$$



$$\begin{array}{l} x = 5.450 \\ u(x) = 0.050 \end{array}$$

or

$$\begin{array}{l} 2u(x) = 0.10016 = 0.10 \\ x = (5.45 \pm 0.10) \end{array}$$

Total uncertainty

Example:

We have performed a series of measurements getting the following results x_1, x_2, \dots, x_n . In such a sample that can be considered as big some of the results are the same; n_k is a number of random experiments, in which the same result x_k has occurred.

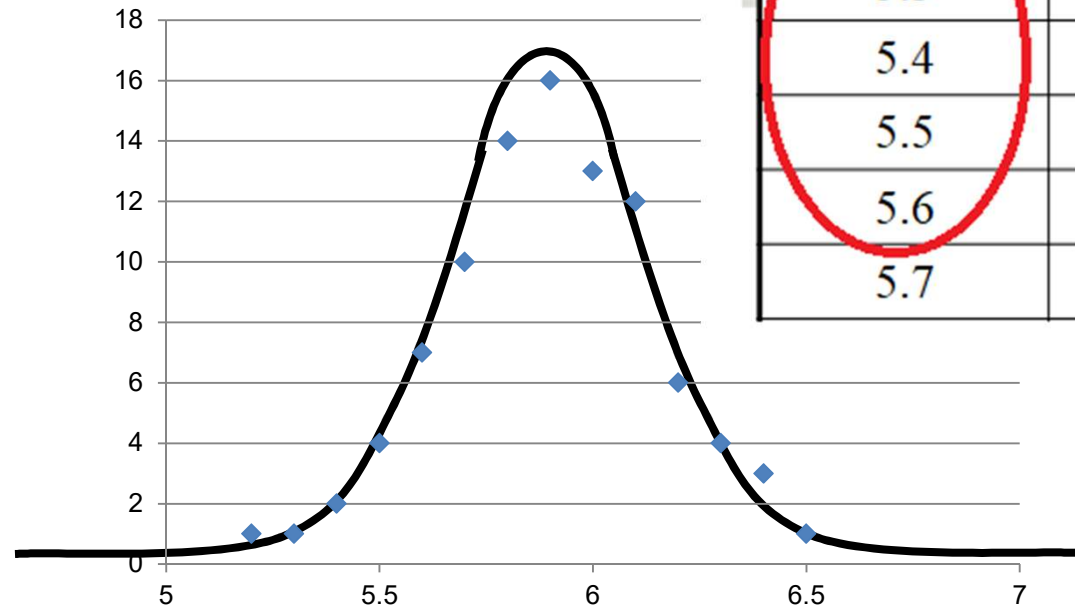
x_k	n_k
5.2	1
5.3	1
5.4	2
5.5	4
5.6	7
5.7	10
5.8	14
5.9	16
6	13
6.1	12
6.2	6
6.3	4
6.4	3
6.5	1
Sum	94

Total uncertainty

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = 5.897872$$

Standard deviation
(of the average):

$$\sigma_{AV} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n(n-1)}} = 0.026328 < 0.1$$



Total uncertainty

Total

$$u_{tot}(x) = \sqrt{u_{random}^2(x) + u_{systematic}^2(x)}$$

$$u_{random}(x) = \sigma_{AV} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n(n-1)}}$$

If the maximum uncertainty is given, it is possible to determine the standard uncertainty:

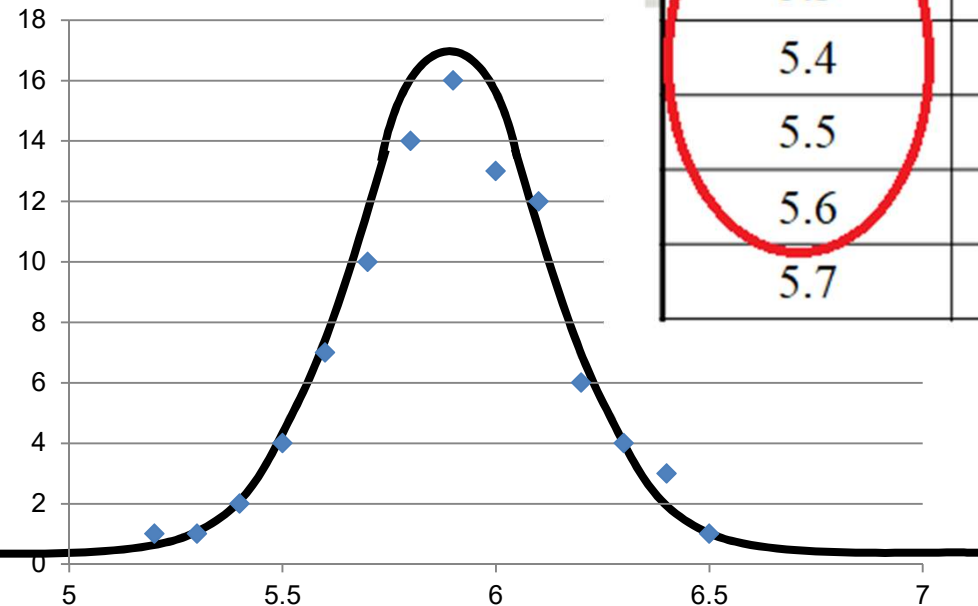
$$u_{systematic}(x) = \frac{\Delta x}{\sqrt{3}}$$

Total uncertainty

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = 5.897872$$

$$\sigma_{AV} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n(n-1)}} = 0.026328$$

$$u_{tot}(x) = \sqrt{0.026328^2 + \left(\frac{0.1}{\sqrt{3}}\right)^2} = 0.063517 = 0.064$$



$$\Delta x = 2 \cdot 0.063517 = 0.127034 = 0.13$$

Answer:

$$x = 5.898, \quad u(x) = 0.064 \quad \text{or}$$

$$x = (5.90 \pm 0.13)$$

Resume

1. Take several measurements.
2. If the results are identical, take the scale of the instrument as the uncertainty - it is the systematic and the maximal uncertainty.

$$x = x \pm \Delta x$$

Resume

3. If the results are different, take at least 5 measurements. Find the estimator of the standard deviation of the mean (average) for these results.

$$\sigma_{AV} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n(n-1)}}$$

Resume

.4. If the standard deviation is greater than the systematic uncertainty (instrument scale), assume that it is the random uncertainty of the measured quantity (neglect the effect of the instrument scale). You can also determine the maximum uncertainty (95% of the results).

$$\sigma_{AV} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n(n-1)}} > \Delta_{\text{systematic}}(x)$$

$$x = \bar{x}$$

$$u(x) = \sigma_{AV}$$

$$\Delta x = 2u(x)$$

$$x = \bar{x} \pm \Delta x$$

Resume

5. If the calculated standard deviation is less than the systematic uncertainty (instrument scale interval), calculate the total standard uncertainty. To calculate the standard total uncertainty, you should change the instrument scale interval (maximum systematic uncertainty) to the standard systematic uncertainty.

$$u_{tot}(x) = \sqrt{u_{random}^2(x) + u_{systematic}^2(x)}$$

$$u_{random}(x) = \sigma_{AV} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n(n-1)}}$$

$$u_{systematic}(x) = \frac{\Delta x}{\sqrt{3}}$$

6. You can extend the total systematic uncertainty to the total maximum uncertainty.

$$\Delta x = 2u(x)$$

$$u(x) = \sqrt{\sigma_{AV}^2 + \left(\frac{\Delta x}{\sqrt{3}}\right)^2}$$

$$x = \bar{x} \pm \Delta x$$

Absolute and relative uncertainty

Absolute uncertainty is expressed in the same units as a measurand

Symbols: $u(x)$ or Δx

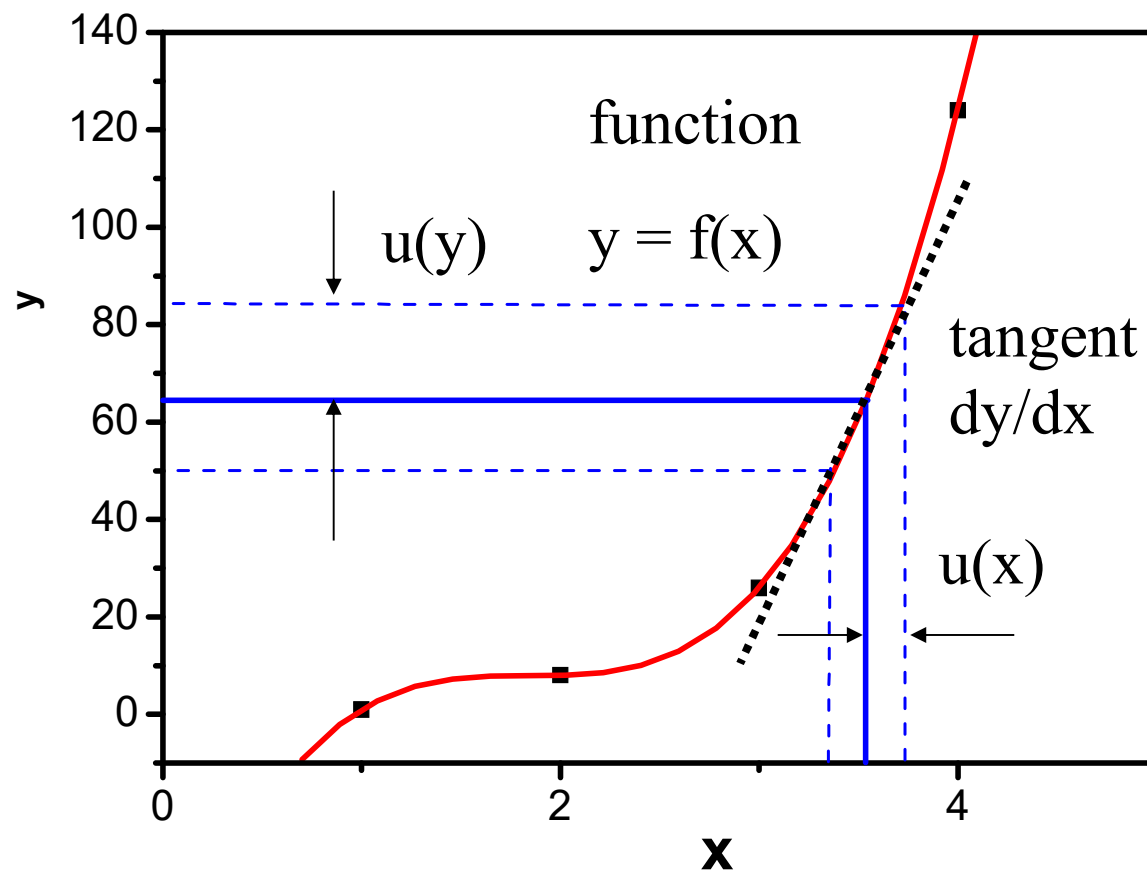
Relative uncertainty $u_r(x)$ or Δx_r the ratio of absolute uncertainty to the measured value:

$$u_r(x) = \frac{u(x)}{x}$$

$$\Delta x_r = \frac{\Delta x}{x}$$

Relative uncertainty has no units and can be expressed in %

Uncertainty of complex measurand – propagation of errors



$$u(y) = \frac{dy}{dx} u(x)$$

Law of propagation of uncertainties

Standard uncertainty of complex measurand $y=f(x_1, x_2, \dots, x_n)$ can be calculated from the law of propagation of uncertainties as a geometric sum of partial differentials.

$$u_c(y) = \sqrt{\left[\frac{\partial y}{\partial x_1} u(x_1) \right]^2 + \left[\frac{\partial y}{\partial x_2} u(x_2) \right]^2 + \dots + \left[\frac{\partial y}{\partial x_n} u(x_n) \right]^2}$$

$$u_{cr}(y) = \frac{u_c(y)}{y}$$

$$g = \frac{4\pi^2 l}{T^2} = 4\pi^2 \cdot l \cdot T^{-2}$$

$$\frac{\partial g}{\partial l} = \frac{4\pi^2}{T^2}$$

$$\frac{\partial g}{\partial T} = -2 \cdot 4\pi^2 \cdot l \cdot T^{-3} = -\frac{8\pi^2 l}{T^3}$$

$$\mu(g) = \sqrt{\left(\frac{\partial g}{\partial l} \cdot u(l)\right)^2 + \left(\frac{\partial g}{\partial T} u(T)\right)^2}$$

$\frac{\Delta l}{\sqrt{3}}$

Example

In a certain experiment one determines gravitational acceleration g on Earth by measuring the period T and length L of a mathematical pendulum. Directly measured length is reported as (1.1325 ± 0.0001) m. The period is $T = 2.12$ s, $u(T) = 0.21$ s. Calculate the relative uncertainty of g .

$$\frac{\partial g}{\partial L} = \frac{4\pi^2}{T^2} = 8,7839 \quad ; \quad u(L) = \frac{0.0001}{\sqrt{3}}$$

$$\frac{\partial g}{\partial T} = \frac{-8\pi^2 L}{T^3} = -9,3847 \quad ; \quad u(T) = 0.21$$

$$\mu(g) = \sqrt{\left(8,7839 \cdot \frac{0,0001}{\sqrt{3}}\right)^2 + \left(3,3847 \cdot 0,21\right)^2} =$$

$$= 1,971 \approx$$

$$= 2,0 \left[\frac{m}{s^2} \right]$$

$$\frac{\partial g}{\partial l} = \frac{4\pi^2}{T^2} = 8,7839 ; \mu(l) = \frac{0,0001}{\sqrt{3}}$$

$$\frac{\partial g}{\partial T} = -\frac{8\pi^2 l}{T^3} = -3,3847 ; \mu(T) = 0,21$$

$$g = \frac{4\pi^2 l}{T^2} = 9,948 \approx 9,9 \left[\frac{m}{s^2} \right]$$

$$g = 9,9 \frac{m}{s^2} ; \mu(g) = 2,0 \frac{m}{s^2}$$

Least Square Method - Linear Regression

Please calculate the density for each measurement point, then determine the mean value and standard deviation of the mean

V [cm ³]	1.0	2.0	3.0	4.0	5.0	6.0	7.0
m [g]	8.1	15.7	26.1	30.3	43.2	48	55.3

Least Square Method - Linear Regression

V [cm ³]	1.0	2.0	3.0	4.0	5.0	6.0	7.0
m [g]	8.1	15.7	26.1	30.3	43.2	48	55.3

$$\rho \left[\frac{\text{g}}{\text{cm}^3} \right] \quad 8.1 \quad 7.85 \quad 8.7 \quad 7.575 \quad 8.64 \quad 8 \quad 7.9$$

$$\bar{\rho} = 8.10329$$

$$\sigma_{Av} = \sqrt{\frac{\sum_{i=1}^7 (\rho_i - \bar{\rho})^2}{7(7-1)}}$$

$$= 0.15729$$

$$= \frac{8.63041 \cdot 10^{-5} + 0.06723 + 0.77893 + 0.29547 + 0.28165 + 0.01194 + 0.0438}{42} =$$

Least Square Method - Linear Regression

V [cm³]	1.0	2.0	3.0	4.0	5.0	6.0	7.0
m [g]	8.1	15.7	26.1	30.3	43.2	48	55.3

$$\rho = \frac{m}{V}$$

$$\Delta m = 0.1$$

$$\Delta V = 0.1$$

$$u(\rho) = \sqrt{\left(\frac{\partial \rho}{\partial m} u(m)\right)^2 + \left(\frac{\partial \rho}{\partial V} u(V)\right)^2}$$

$$= \sqrt{\left(\frac{1}{V} u(m)\right)^2 + \left(-\frac{m}{V^2} u(V)\right)^2} = 0.1685...$$

$\frac{1}{V} \underset{\substack{0.1 \\ 13}}{\quad}$
 $-\frac{m}{V^2} \underset{\substack{0.1 \\ 13}}{\quad}$

$$\mu(\rho)_{\text{Tot}} = \sqrt{\sigma_{Av}^2 + \mu(\rho)^2} \approx \sqrt{0.15^2 + 0.16^2} \approx 0.2193$$

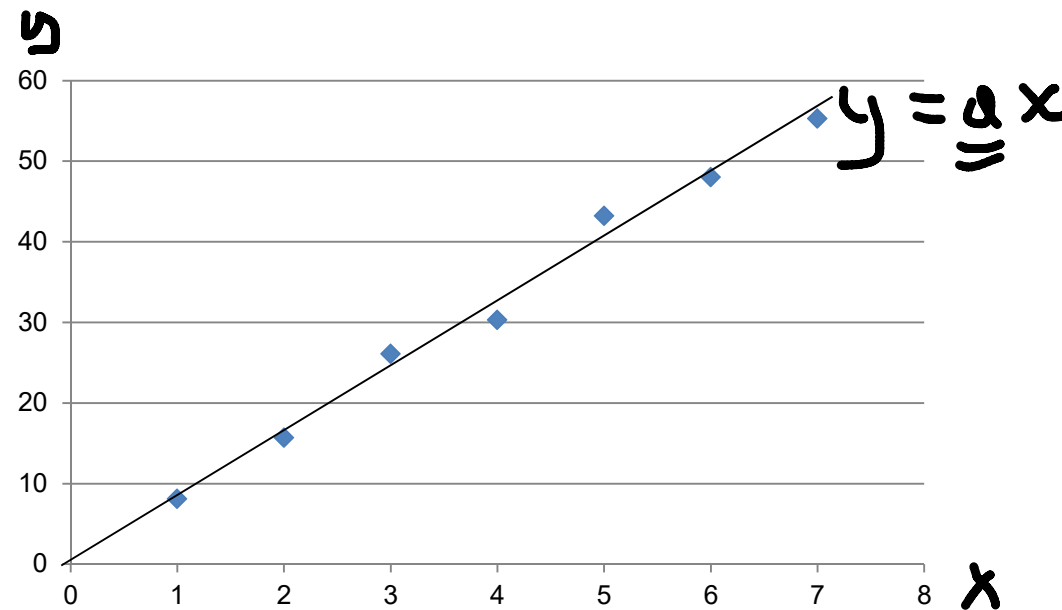
$$\rho = 8.11 \text{ g/cm}^3$$

$$\mu(\rho) = 0.22 \text{ g/cm}^3$$

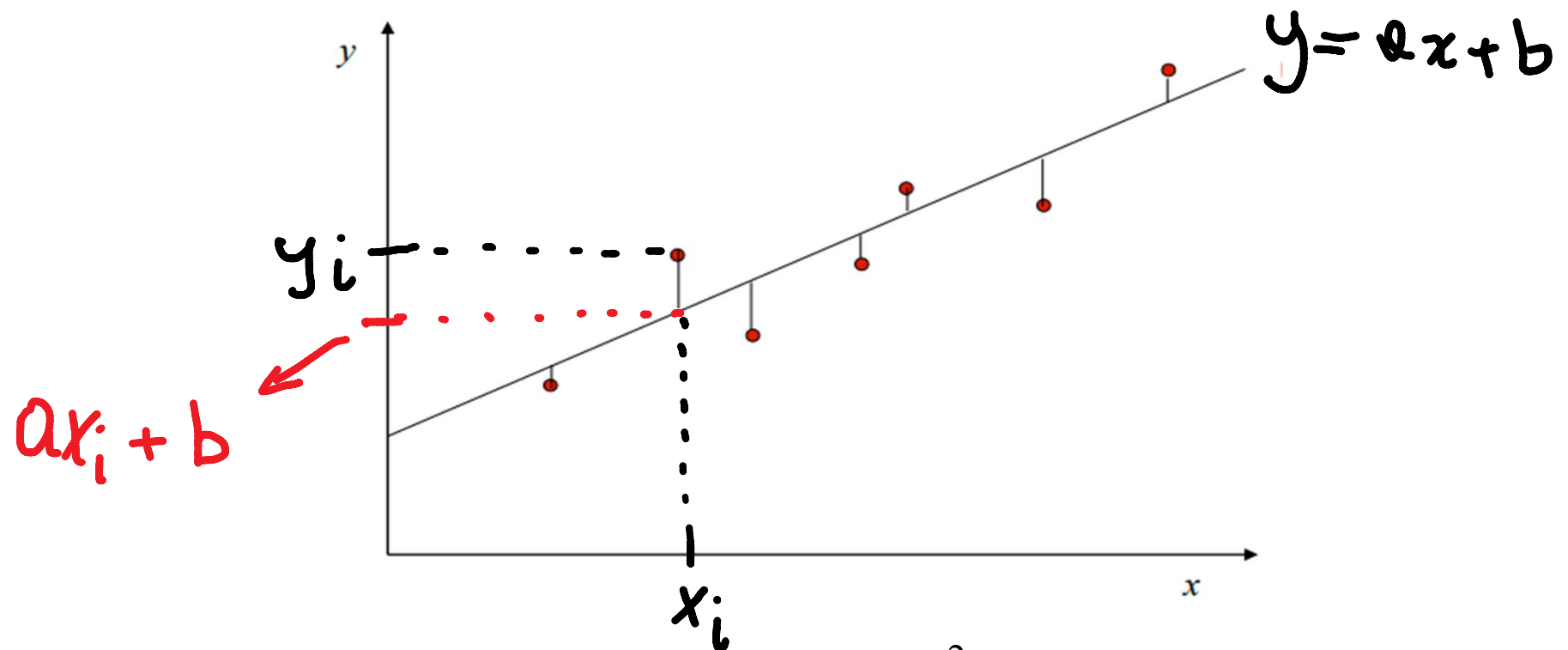
Least Square Method - Linear Regression

V [cm ³]	1.0	2.0	3.0	4.0	5.0	6.0	7.0
m [g]	8.1	15.7	26.1	30.3	43.2	48	55.3

$$m = \rho \cdot V$$



Least Square Method - Linear Regression



$$S^2 = \sum_i^n [y_i - (ax_i + b)]^2 = \min$$

Least Square Method - Linear Regression

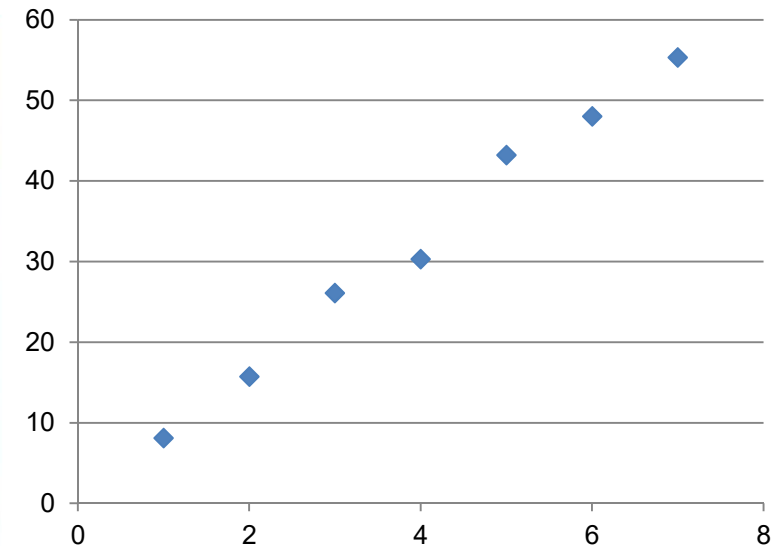
$$S^2 = \sum_i^n [y_i - ax_i]^2 = \min$$

$$\frac{\partial S^2}{\partial a} = 0$$

$$-2 \sum x_i y_i + 2a \sum x_i^2 = 0$$

$$a = \frac{\sum x_i y_i}{\sum x_i^2}$$

$$\sigma_a = \sqrt{\frac{1}{n-1} \cdot \frac{\sum_{i=1}^n [(ax_i - y_i)^2]}{\sum_{i=1}^n [x_i^2]}}$$



$$x = \sum_i^n [y_i - ax_i]^2 = \min$$

$$f(a)$$

$$\frac{\partial S^2}{\partial a} = 0$$

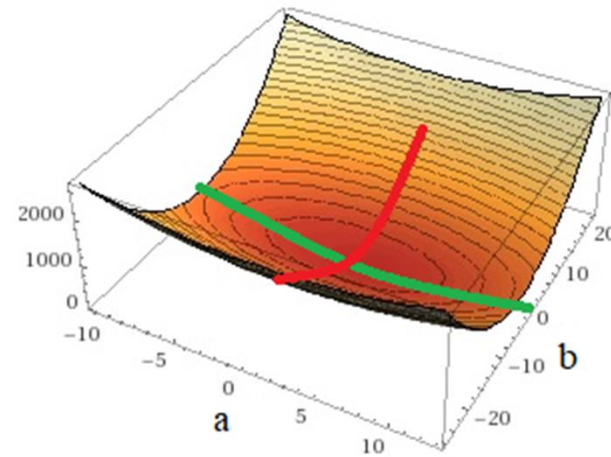
$$f(a) = (y_1 - ax_1)^2 + (y_2 - ax_2)^2 + \dots + (y_n - ax_n)^2$$

$$\frac{\partial f}{\partial a} = 2(y_1 - ax_1) \cdot (-x_1) + 2(y_2 - ax_2) \cdot (-x_2) + \dots =$$

$$= -2 \sum_{i=1}^n x_i (y_i - ax_i) = -2 \sum_{i=1}^n x_i y_i + 2a \sum_{i=1}^n x_i^2$$

Least Square Method - Linear Regression

$$f(a,b) = \sum_{i=1}^n (y_i - ax_i - b)^2 = \min$$



$$\frac{\partial f(a,b)}{\partial a} = 0$$

$$\frac{\partial f(a,b)}{\partial b} = 0$$

Least Square Method - Linear Regression

$$f(a,b) = \sum_{i=1}^n (y_i - ax_i - b)^2 = \min$$

$$-2 \sum_{i=1}^n x_i (y_i - ax_i - b) = 0$$

$$-2 \sum_{i=1}^n (y_i - ax_i - b) = 0$$

$$f(a,b) = \sum_{i=1}^n (y_i - ax_i - b)^2 = \min \quad f(a,b) = (y_1 - ax_1 - b)^2 + (y_2 - ax_2 - b)^2$$

$$-2 \sum_{i=1}^n x_i (y_i - ax_i - b) = 0 \quad + \dots + (y_n - ax_n - b)^2$$

$$-2 \sum_{i=1}^n (y_i - ax_i - b) = 0$$

$$\frac{\partial f}{\partial a} = 2(y_1 - ax_1 - b)(-x_1) + 2(y_2 - ax_2 - b)(-x_2) +$$

$$+ \dots + 2(y_n - ax_n - b)(-x_n)$$

$$\frac{\partial f}{\partial b} = -2(y_1 - ax_1 - b) - 2(y_2 - ax_2 - b) - \dots - 2(y_n - ax_n - b)$$

$$+ \dots + (y_n - ax_n - b)^2$$

$$\frac{\partial F}{\partial a} = 2(y_1 - ax_1 - b)(-x_1) + 2(y_2 - ax_2 - b)(-x_2) + \dots + 2(y_n - ax_n - b)(-x_n)$$

$$\frac{\partial F}{\partial b} = -2(y_1 - ax_1 - b) - 2(y_2 - ax_2 - b) - \dots - 2(y_n - ax_n - b)$$

$$\begin{cases} (y_1 - ax_1 - b)x_1 + (y_2 - ax_2 - b)x_2 + \dots = 0 \\ y_1 - ax_1 - \underline{b} + y_2 - ax_2 - \underline{b} + \dots = 0 \quad / + n \cdot b \quad / : n \end{cases}$$

$$\begin{cases} y_1 x_1 - ax_1^2 - bx_1 + y_2 x_2 - ax_2^2 - bx_2 + \dots = 0 \\ \underline{y_1 - ax_1 + y_2 - ax_2 + \dots} = b \end{cases}$$

$$\begin{cases} (y_1 - ax_1 - b)x_1 + (y_2 - ax_2 - b)x_2 + \dots = 0 \\ y_1 - ax_1 - b + y_2 - ax_2 - b + \dots = 0 \quad | :n \end{cases}$$

AGH

$$\begin{cases} y_1 x_1 - ax_1^2 - bx_1 + y_2 x_2 - ax_2^2 - bx_2 + \dots = 0 \\ y_1 - ax_1 + y_2 - ax_2 + \dots = b \end{cases}$$

$$\sum_{i=1}^n x_i y_i - a \sum_{i=1}^n x_i^2 - \frac{\sum_{i=1}^n y_i - a \sum_{i=1}^n x_i}{\sum_{i=1}^n 1} \cdot \sum_{i=1}^n x_i = 0$$

$$n \sum_{i=1}^n x_i y_i - a n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n y_i + a \sum_{i=1}^n x_i \right) \sum_{i=1}^n x_i = 0$$

$$n \sum_{i=1}^n x_i y_i - a n \sum_{i=1}^n x_i^2 - \sum_{i=1}^n y_i \cdot \sum_{i=1}^n x_i + a \left(\sum_{i=1}^n x_i \right)^2 = 0$$

$$n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n y_i \cdot \sum_{i=1}^n x_i = a \left(n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 \right)$$

$$Q = \frac{n \sum x_i y_i - \sum x \cdot \sum y}{n \sum x_i^2 - \left(\sum x_i \right)^2}$$

Least Square Method - Linear Regression

$$f(a, b) = \sum_{i=1}^n (y_i - ax_i - b)^2 = \min$$

$$a = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2}$$

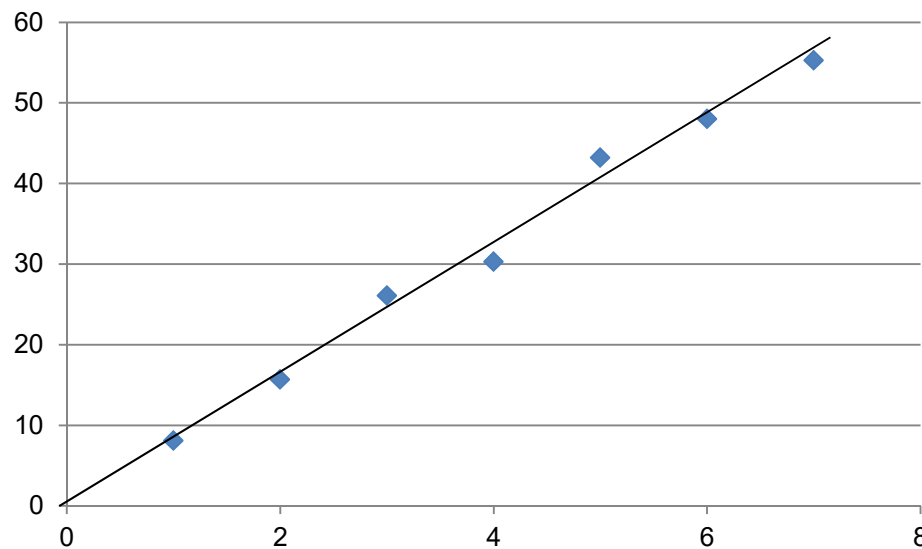
$$u(a) = \sqrt{\frac{n \left[\sum_{i=1}^n y_i^2 - \bar{a} \sum_{i=1}^n x_i y_i - \bar{b} \sum_{i=1}^n y_i \right]}{(n-2) \left[n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 \right]}}$$

$$b = \frac{1}{n} \left(\sum_{i=1}^n y_i - a \sum_{i=1}^n x_i \right)$$

$$u(b) = u(a) \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}$$

Least Square Method - Linear Regression

x	V [cm ³]	1.0	2.0	3.0	4.0	5.0	6.0	7.0
y	m [g]	8.1	15.7	26.1	30.3	43.2	48	55.3



$$d = a = \frac{\sum_{i=1}^7 V_i \cdot m_i}{\sum_{i=1}^7 V_i^2}$$

$$a = \frac{\sum x_i \cdot y_i}{\sum x_i^2}$$

x	V [cm ³]	1.0	2.0	3.0	4.0	5.0	6.0	7.0
y	m [g]	8.1	15.7	26.1	30.3	43.2	48	55.3

$$\sigma_a = \sqrt{\frac{1}{n-1} \cdot \frac{\sum_{i=1}^n [(a_i - \bar{a})^2]}{\sum_{i=1}^n 1}}$$

$$\bar{y} = \frac{8.1 \cdot 1 + 2 \cdot 15.7 + \dots}{1^2 + 2^2 + \dots} = 8.0721$$

$$\begin{aligned} \mu(y) &= \sqrt{\frac{1}{7-1} \cdot \frac{(8.0721 \cdot 1 - 8.1)^2 + (8.0721 \cdot 2 - 15.7)^2 + \dots}{1^2 + 2^2 + \dots}} \\ &= \sqrt{\frac{1}{6} \cdot \frac{7.78 \cdot 10^{-4} + 0.1913 + 3.5483 + 3.9537 + 8.0628 + 0.1871}{140} + 1.4513} = 0.1439 \end{aligned}$$

$$\rho = 8.07 \text{ g/cm}^3$$

$$\mu(\rho) = 0.15 \text{ g/cm}^3$$

$$\Delta \rho = 2 \cdot \mu(\rho) = 0.1439 \cdot 2 = 0.2878 \approx 0.29$$

$$\rho = (8.07 \pm 0.29) \text{ g/cm}^3$$

$$a = \frac{4\pi^2}{g}$$

$$g = \frac{4\pi^2}{a}$$

$$T = 2\pi \sqrt{l/g}$$

$$g = \frac{4\pi^2}{T^2} \cdot l$$

$$\Rightarrow T^2 = \frac{4\pi^2}{g} \cdot l$$
