Error calculation

(+Examples)

Introduction for students of the TUHH

Introduction

My name is Wladimir Banda-Barragán.

I will be your tutor in the first two laboratory sessions.

The class will be organised as follows:

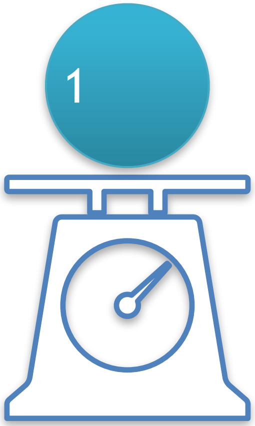
1. Lecture with step-by-step examples (1.5 hours).
2. Exercises in the protocol (2 hours)
3. Questions and wrap-up (0.5 hours)

Measuring

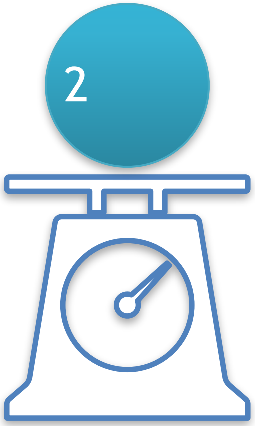
Experimental physics is all about measuring

There is no exact measurement!

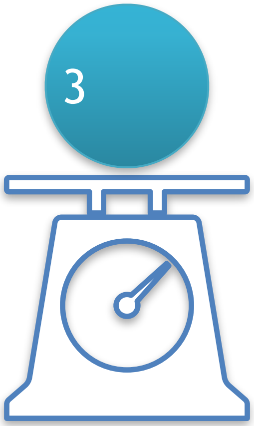
Measurements spread unavoidably



$m_1 = 78.34655\,\mathrm{kg}$



$m_2 = 78.55089\,\mathrm{kg}$



$m_3 = 78.49862\,\mathrm{kg}$

Measuring

But the measurements tend to accumulate near one value

$$m_1 = 78.34655\,\mathrm{kg}$$

$$m_2 = 78.55089\,\mathrm{kg}$$

$$m_3 = 78.49862\,\mathrm{kg}$$

$$\bar{m} = 78.46535\,\mathrm{kg}$$

Uncertainties (also called errors) reflect this dispersion

In day-to-day life one (experienced) measurement is enough if the **spread** is smaller than the **accuracy** condition

E.g. a wooden board measured once by a carpenter usually fits

Precision and accuracy

To understand the difference between these two concepts, let us use this applet:

We will study the accelerated motion of a massless vehicle, initially at rest:

https://www.walter-fendt.de/html5/phen/acceleration_en.htm

And let us get a mobile phone chronometer or an online one, e.g., see:

http://online-stopwatch.chronme.com/

Precision and accuracy

We choose an initial acceleration of

$$a = +2\frac{m}{s^2}$$

From Newton's laws we know that:

$$t = \sqrt{\frac{2d}{a}}$$

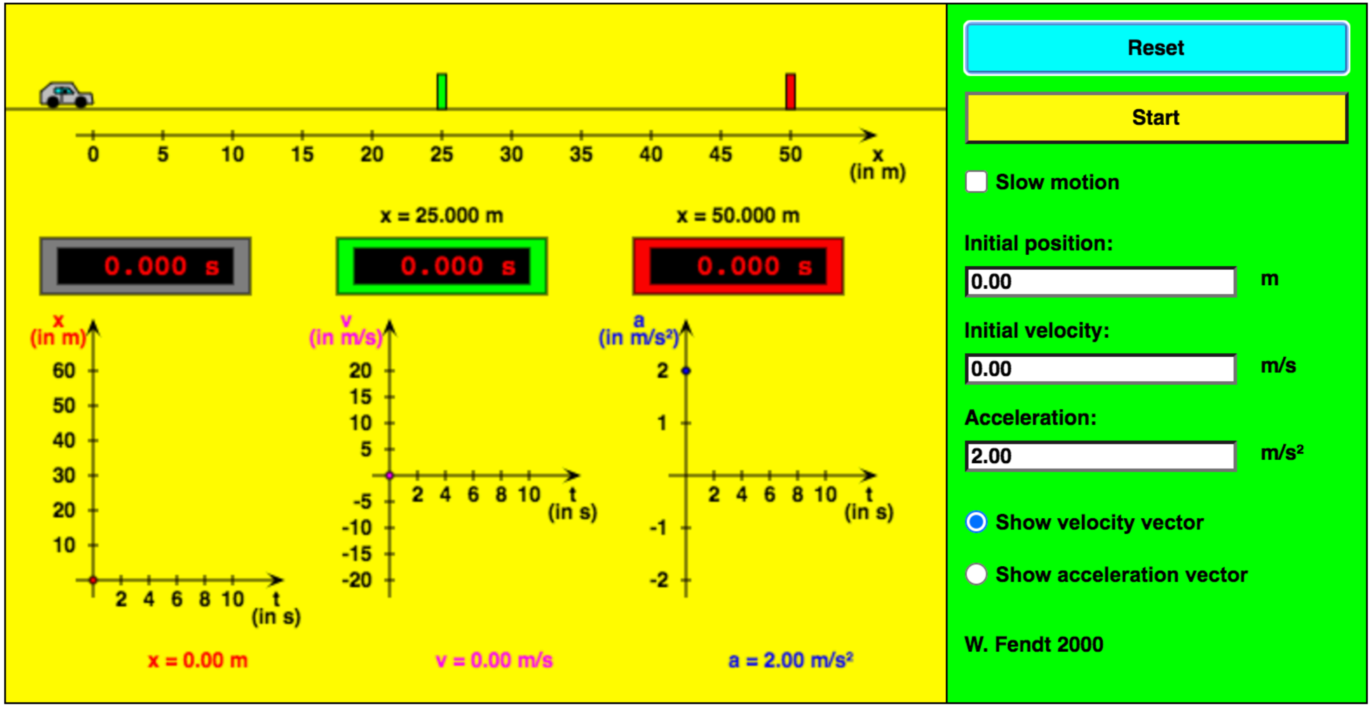
If $d = 25m \Rightarrow t_{real} = 5.00s$

Calibrated

Not calibrated

N	t_{good}	t_{bad}
1	5.09	5.33
2	4.03	5.45
3	4.99	5.36
4	5.33	5.46
5	5.03	5.50
Average	4.894	5.42

Precision and accuracy



Calibrated

t_{good}

Not calibrated

t_{bad}

Precision and accuracy

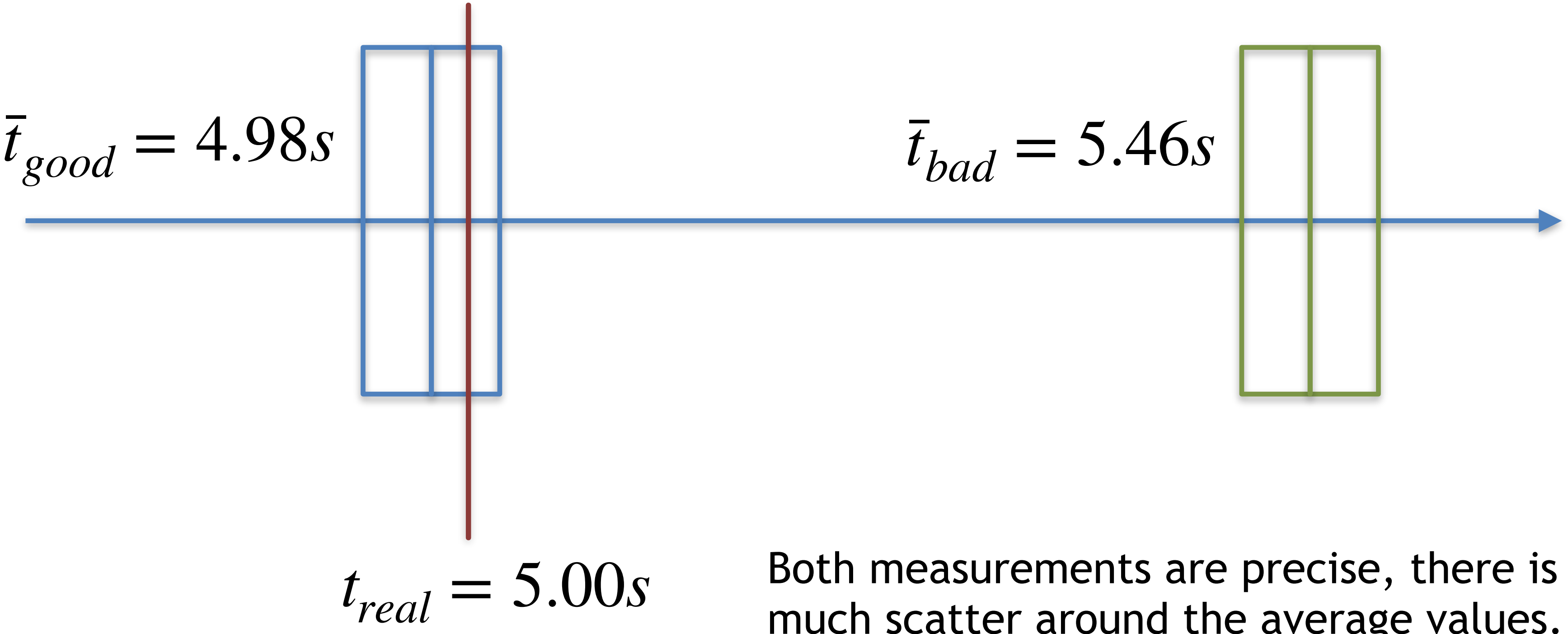
$$t_{real} = 5.00s$$

$$t_{good} = 4.98s$$ Calibrated

$$t_{bad} = 5.46s$$ Not calibrated

	Calibrated	Not calibrated
N	t_{good}	t_{bad}
1	5.10	5.49
2	4.90	5.35
3	5.05	5.52
4	4.76	5.50
5	5.09	5.44
Average	**4.98**	**5.46**

Precision and accuracy



Both measurements are precise, there is no much scatter around the average values.

However, only one is accurate.

Precision and accuracy

Precision is a description of random errors, a measure of statistical variability.

Accuracy has two definitions:

1. More commonly, it is a description of systematic errors, a measure of statistical bias; low accuracy causes a difference between a result and a "true" value. ISO calls this trueness.

2. Alternatively, ISO defines[1] accuracy as describing a combination of both types of observational error above (random and systematic), so high accuracy requires both high precision and high trueness.

Source: Wikipedia

Uncertainties/Errors

In physics we need to determine the error in our measurements

There are 2 kinds of errors:

Statistical and systematic errors

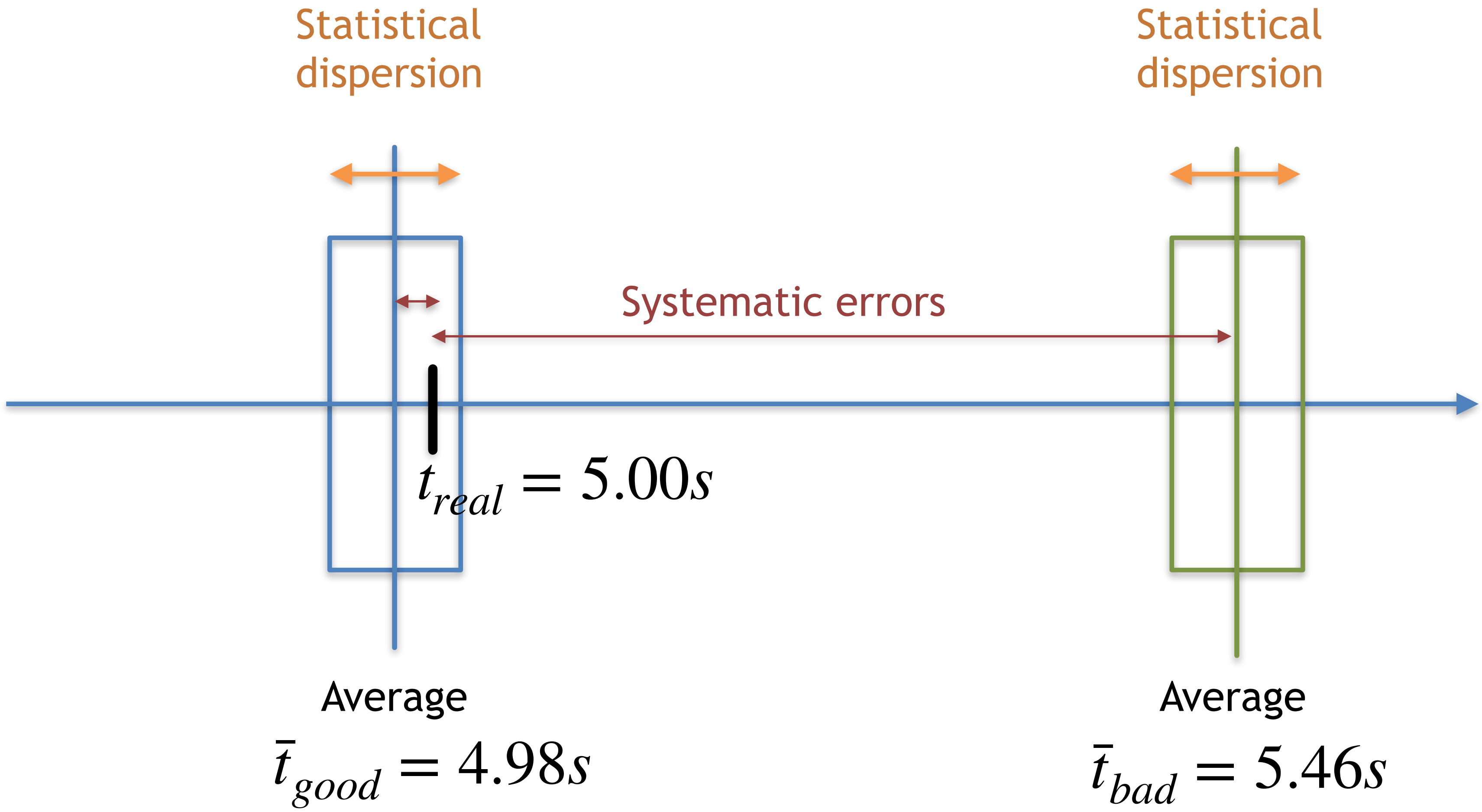
Statistical errors cannot be avoided:

Every measurement process has a limited precision caused by an inherent scatter of the values

Systematic errors should be avoided or at least minimised

Systematic errors may affect the precision and the accuracy

Uncertainties/Errors



Statistical errors

The computation of the statistical error is based on the theory of probability

There is a non-zero probability for every value

Statistical errors

There has to be an infinite number of measurements to determine the true value, the mean of the population

Real measurements can only be a sample of the population of measurements

The mean value of a sample **is only an estimate** of the true value

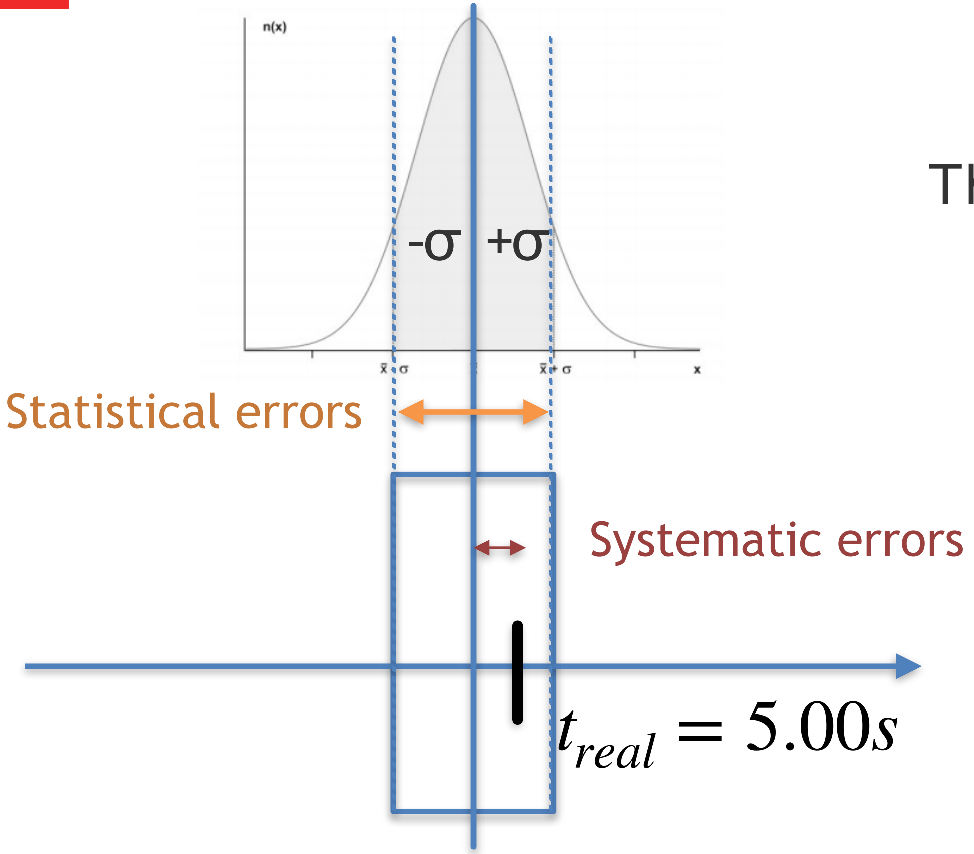
Systematic errors

Systematic errors affect all the individual measurements in the same way.

They are reproducible, i.e. they occur in the same size and with the same sign by repeating the measurements under the same conditions.

The results of measurement can and must be corrected accordingly. Note that the statistical errors are not reproducible!

Frequency distribution

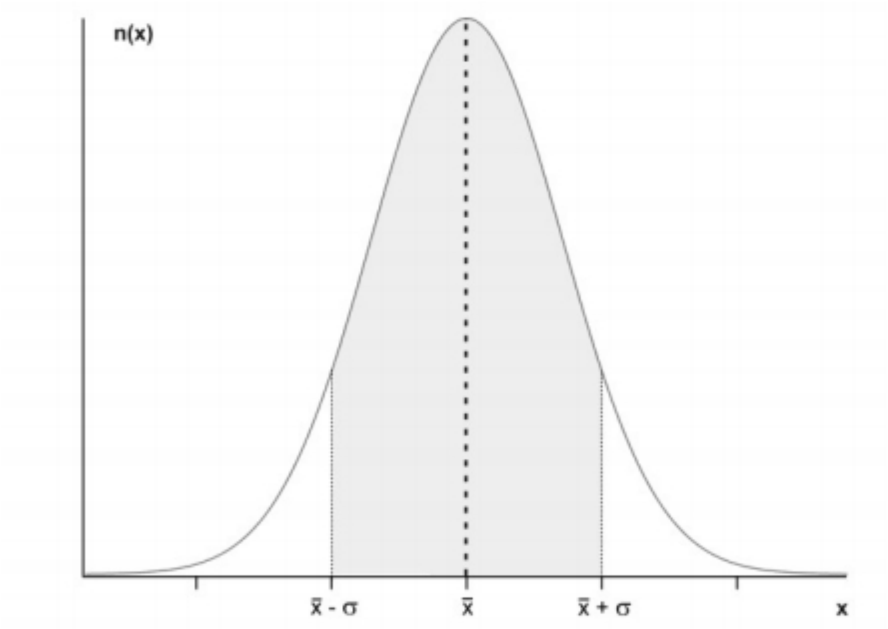


Average $\bar{t}_{good} = 4.98s$

The frequency of the values of the population are described by a Gaussian (Normal) distribution

Frequency distribution

The frequency of the values of the population are described by a Gaussian (Normal) distribution



$$n(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}}$$

$$\int_{-\infty}^{+\infty} n(x)\,dx = 1$$

$$\int_{\bar{x}-\sigma}^{\bar{x}+\sigma} n(x)\,dx = 0.683$$

Gaussian (Normal) distribution

$$n(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}}$$

n(x) is the distribution of a probability density

n(x)dx is the probability of a value being in [x, x + dx]

$\bar{x}$ is the mean (most probable) value of the distribution

σ is an indicator for the spread (precision) of the population

Sample of measurements

Sample of measurements allows an estimate of the true value

Best estimate of the true value for N measurements:

Mean value:

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

Standard deviation:

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

Mean and standard deviation

N	$t_{good} \equiv x$	$t_{bad} \equiv x$
1	5.10	5.49
2	4.90	5.35
3	5.05	5.52
4	4.76	5.50
5	5.09	5.44
Average		
Standard deviation		

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

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

Mean and standard deviation

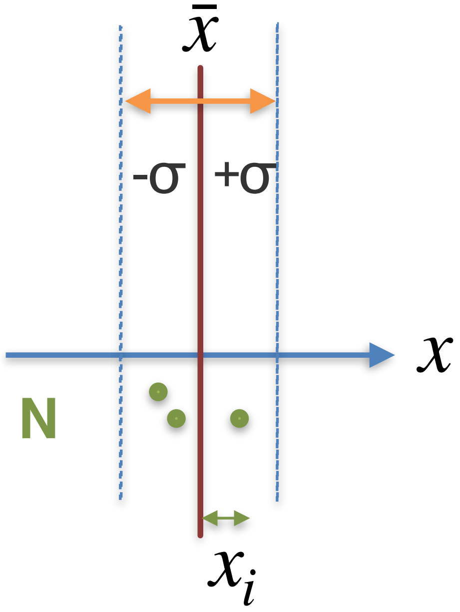
N	t_{good} (s)	t_{bad} (s)
1	5.10	5.49
2	4.90	5.35
3	5.05	5.52
4	4.76	5.50
5	5.09	5.44
Average	**4.98**	**5.46**
Standard deviation	**0.15**	**0.07**

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

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

Standard deviation

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

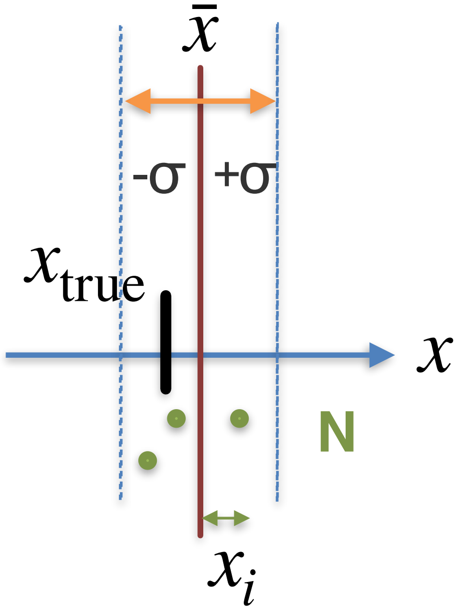


σ_x is the "error" (precision) of the estimate

Mean distance of measurements to their average

Denominator of N-1 instead of N reduces the bias due to the finite sample size (it is called Bessel's correction)

Standard deviation

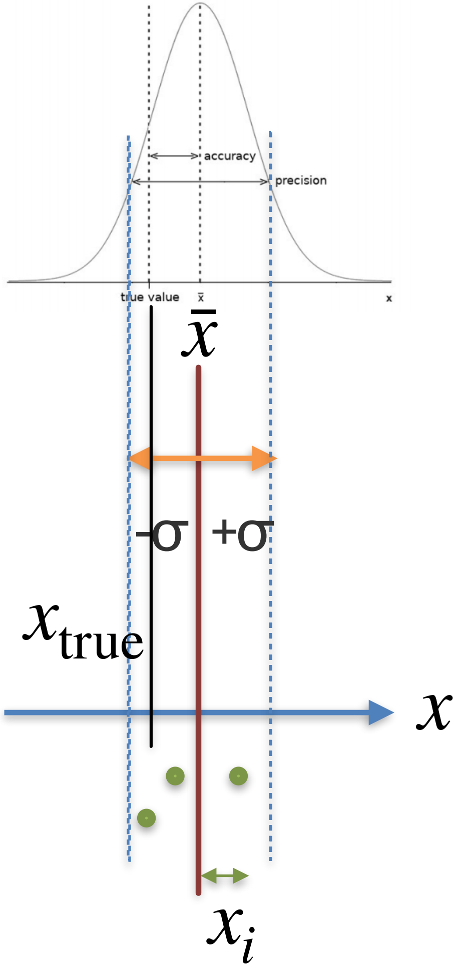


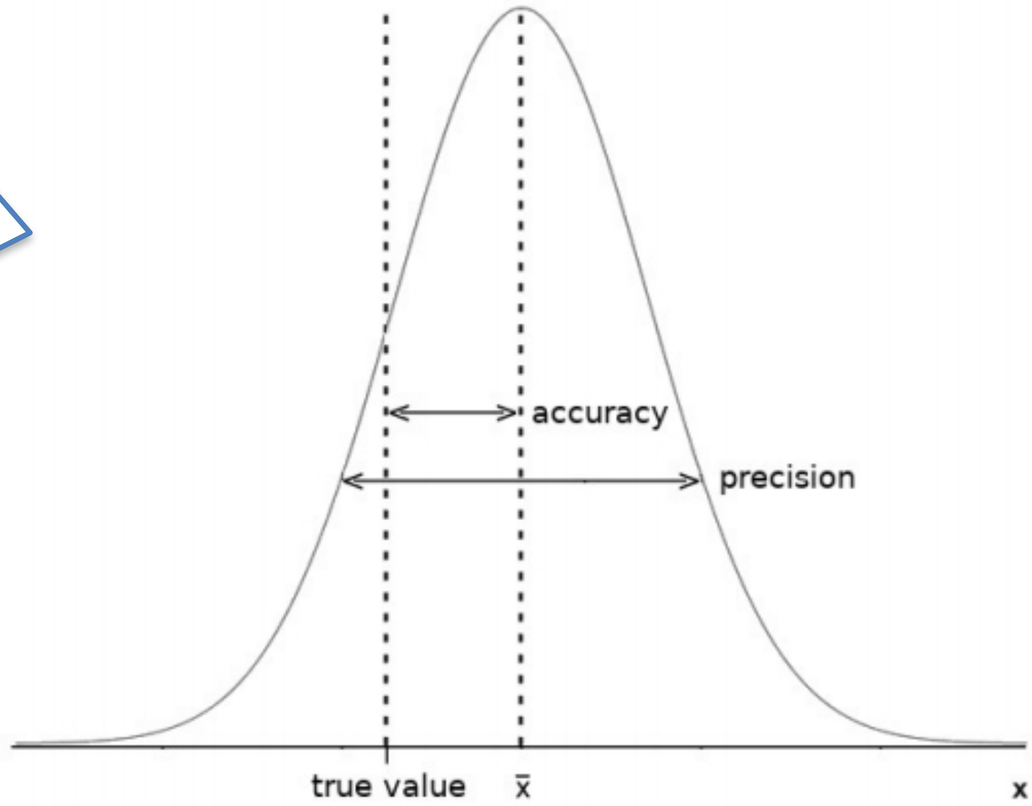
To estimate the accuracy of $\bar{x}$ we can use:

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

So, with increasing **N**, the estimate approaches the true value

Accuracy and precision of sample





Specifying quantities with errors

Quantity m, error Δm

$m \pm \Delta m$

Number of decimals depends on size of Δm

e.g. $m=13.75643$, $\Delta m=0.4378$

$m=13.8 \pm 0.4$ (significant first decimal of error)

e.g. $m=13.75643$, $\Delta m=0.1378$

$m=13.76 \pm 0.14$ (significant first two, if first is 1)

Specifying quantities with errors

N	x (s)
1	5.10
2	4.90
3	5.05
4	4.76
5	5.09
Average	**4.980**
Standard deviation	**0.147**

In our example:

$$\bar{x} = 4.980 \pm 0.147\ s$$

$$\bar{x} = 4.98 \pm 0.15\ s$$

$$\bar{x} = 5.0 \pm 0.2\ s$$

Error of compound quantities

Physical quantity deduced from different measured values.

e.g. density from mass and volume, or the acceleration of gravity from height “h” and time “t” (free-fall experiment).

$$\rho = \frac{m}{V} \qquad\qquad g = \frac{2h}{t^2}$$

For example, we can estimate the error in “g” from the errors in “h” and “t”, which we can measure experimentally.

Error of compound quantities

Error is computable if:

The **measured values are normally distributed** with statistically determined errors

The **measured values are independent**, thus not deductive from each other

Error propagation

Let the quantity f be function of n measured values p_i $i \in \{1,\dots,n\}$

$$f\left(p_1, p_2, \dots, p_n\right)$$

The error σ_f of f() :

$$\sigma_f = \sqrt{\sum_{i=1}^{n} \left(\frac{\delta f}{\delta p_i} \sigma_{p_i} \right)^2}$$

Where:

σ_{p_i} error of p_i

$\frac{\delta f}{\delta p_i}$ partial derivative

Error propagation: Example

Let us work with the first example:

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

$$\sigma_f = \sqrt{\sum_{i=1}^{n} \left(\frac{\delta f}{\delta p_i} \sigma_{p_i} \right)^2}$$

Let us assume we measured m and V independently:

$$m = \bar{m} \pm \sigma_m = 49.45\,\mathrm{kg} \pm 0.74\,\mathrm{kg}$$

$$V = \bar{V} \pm \sigma_V = 3.14\,\mathrm{m}^3 \pm 0.18\,\mathrm{m}^3$$

$$\rho = \frac{m}{V} = 15.75 \frac{\mathrm{kg}}{\mathrm{m}^3}$$

And we obtained the mean density, what is the error σ_ρ?

Error propagation: Example

Let us work with the first example:

$$\rho = \rho(m, V) = \frac{m}{V}$$

$$\sigma_f = \sqrt{\sum_{i=1}^{n} \left(\frac{\delta f}{\delta p_i} \sigma_{p_i} \right)^2}$$

What is the error σ_ρ?

$$\sigma_\rho = \sqrt{\left(\frac{\delta \rho}{\delta m} \sigma_m \right)^2 + \left(\frac{\delta \rho}{\delta V} \sigma_V \right)^2}$$

$$\sigma_\rho = \rho \sqrt{\left(\frac{\sigma_m}{m} \right)^2 + \left(\frac{\sigma_V}{V} \right)^2} = \pm 0.93 \frac{\mathrm{kg}}{\mathrm{m}^3}$$

$$\rho = \bar{\rho} \pm \sigma_\rho = 15.75 \frac{\mathrm{kg}}{\mathrm{m}^3} \pm 0.93 \frac{\mathrm{kg}}{\mathrm{m}^3}$$

Error propagation: Summary

In summary the errors from the independent measurements of mass and volume propagate to the error of the density.

$$\begin{matrix} \sigma_m \\ \sigma_V \end{matrix} \quad \Longrightarrow \quad \sigma_\rho$$

Some error propagation formulae can be found here:

https://en.wikipedia.org/wiki/Propagation_of_uncertainty

Maximum error

If conditions for the propagation law are not met, we can estimate the maximum error:

$$\sigma_f = \sum_{i=1}^{n} \left| \frac{\delta f}{\delta p_i} \sigma_{p_i} \right|$$

σ_{p_i} may be statistical or estimated errors

Maximum error

In the same example, let us assume one of the criteria was not met.

$$\rho = \rho(m, V) = \frac{m}{V}$$

What is the error σ_ρ?

$$\sigma_f = \sum_{i=1}^{n} \left| \frac{\delta f}{\delta p_i} \sigma_{p_i} \right|$$

$$\sigma_\rho = \left| \frac{\delta \rho}{\delta m} \sigma_m \right| + \left| \frac{\delta \rho}{\delta V} \sigma_V \right|$$

$$\sigma_\rho = \rho \left(\left| \frac{\sigma_m}{m} \right| + \left| \frac{\sigma_V}{V} \right| \right) = \pm 1.14 \frac{\mathrm{kg}}{\mathrm{m}^3}$$

$$\rho = \bar{\rho} \pm \sigma_\rho = 15.75 \frac{\mathrm{kg}}{\mathrm{m}^3} \pm 1.14 \frac{\mathrm{kg}}{\mathrm{m}^3}$$

Linear Regression

Let us assume we have two quantities "x" and "y"

We vary quantity "x" and measure the dependent quantity "y", "n" times in a sequence.

At the end, we have "n" quantity pairs.

i	x_i	y_i
1		
2		
3		
4		
5		
…		
n		

Linear Regression

Is there a linear relation between “x” and “y”

It is always a good idea to make a plot of x vs. y.

Not all relations between x and y will be linear. If they are not, you can **linearise** the relation first.

In either case:

To determine the linear dependency between “x” and “y”, we carry out a **linear regression**

Linear dependent quantities

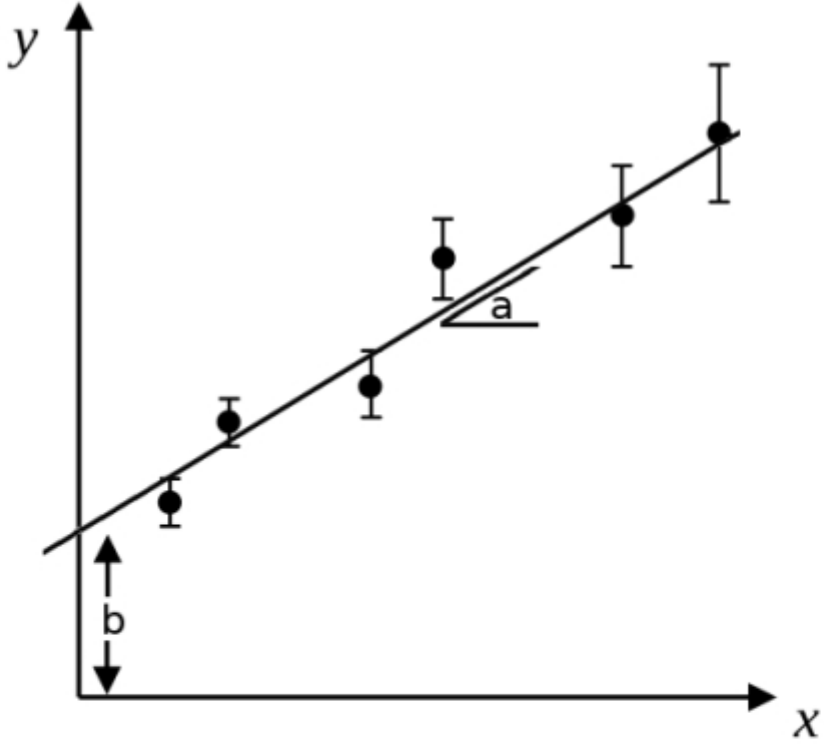
Remember that:

$$y = a \cdot x + b$$

$a : slope$

$b : axis\ intercept$

We need to find “a” and “b” for a straight line that fits best the data points.



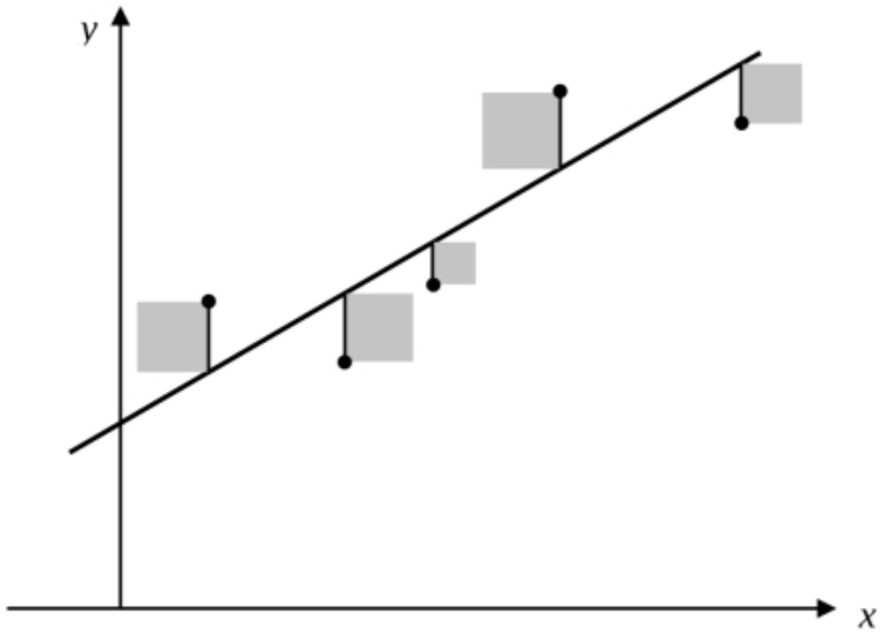
Background: least mean square

We need to minimise this expression (i.e. get the smallest sum of the grey areas)

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

by varying “a” and “b”

(partial derivatives set to 0)



Formulae I: solution

The best “a” and “b” are

$$a = \frac{\sum_{i=1}^{n} \left(x_i - \overline{x}\right) \cdot \left(y_i - \overline{y}\right)}{\sum_{i=1}^{n} \left(x_i - \overline{x}\right)^2} \qquad b = \overline{y} - a \cdot \overline{x}$$

with $\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ and $\overline{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$

Formulae II: errors

Error of the fit

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

Denominator (n-2) as we are fitting 2 parameters

Errors for “a” and “b”

$$\sigma_a = \sigma_y \cdot \sqrt{\frac{1}{\sum_{i=1}^{n} (x_i - \overline{x})^2}}$$

$$\sigma_b = \sigma_y \cdot \sqrt{\frac{1}{n} \cdot \frac{\sum_{i=1}^{n} x_i^2}{\sum_{i=1}^{n} (x_i - \overline{x})^2}}$$

Linear regression: example

Determine the thickness of pages in a book

Measure the thickness "d" of different number "m" of pages (including the cover)

$$d_i = a \cdot m_i + b$$

The slope is the thickness of a page, and the axis intercept is the thickness of the cover.

m	d (mm)
10	3.2
20	4.2
30	5.1
40	5.8
50	6.8
60	7.7
70	8.8
80	9.7
90	10.8
100	11.7

Linear regression: example

m	d (mm)
10	3.2
20	4.2
30	5.1
40	5.8
50	6.8
60	7.7
70	8.8
80	9.7
90	10.8
100	11.7

1. Make a plot and devise a model: $d_i = a \cdot m_i + b$

2. Calculate averages of “m” and “d”.

3. Calculate the slope and intercept with:

$$a = \frac{\sum_{i=1}^{n} (x_i - \overline{x}) \cdot (y_i - \overline{y})}{\sum_{i=1}^{n} (x_i - \overline{x})^2} \qquad b = \overline{y} - a \cdot \overline{x}$$

$\Rightarrow$

a=0.0943 mm

b=2.1935 mm

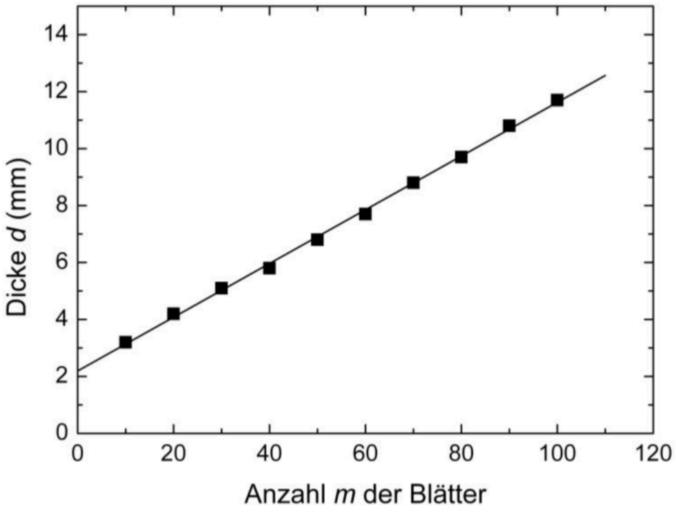
Linear regression: example

m	d (mm)
10	3.2
20	4.2
30	5.1
40	5.8
50	6.8
60	7.7
70	8.8
80	9.7
90	10.8
100	11.7

$$a=\frac{\sum_{i=1}^{n}(x_i-\bar{x})\cdot(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \qquad b=\bar{y}-a\cdot\bar{x}$$

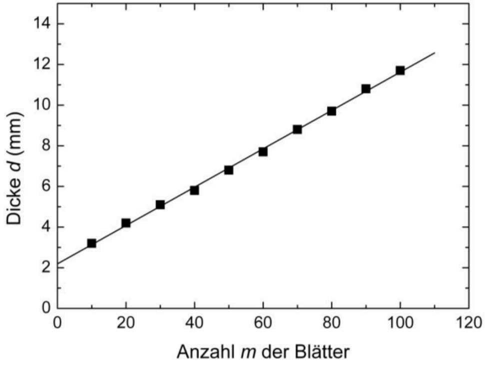
a=0.0943 mm

b=2.1935 mm



d=0.0943 m + 2,1935 [mm]

Linear regression: example



$d=0.0943\ m + 2{,}1935$ [mm]

4. Calculate the errors for the slope and intercept with our equations in slide 40:

$\sigma_d = 0.173$ mm, $\sigma_a = 0.0019$ mm, and $\sigma_b = 0.12$ mm.

5. We report "a" and "b" with their errors:

$a=(0.0943 \pm 0.0019)$ mm (average thickness of a sheet in the book)

$b=(2.19 \pm 0.12)$ mm (thickness of the book back cover)

Book thickness $d = a \cdot m + 2b$

Protocol 1

Tasks:

a. What is the thickness of a single page? What is the thickness of the cover?

b. What is the volume of a single page? What is the volume of 100 pages?

All results must be given with the corresponding errors

Pick up any book with a hard cover, and use a ruler or this online Vernier:

https://www.stefanelli.eng.br/en/virtual-vernier-caliper-simulator-05-millimeter/#swiffycontainer_2

Measurement to be conducted

Number and thickness of the pages

a.

m, Number of the pages (one page has front and back)	25	50	75	100	125
Thickness of the cover and m pages [mm]					

Width and length of a single page

b.

	1. measurement	**2. measurement**	**3. measurement**	**4. measurement**	**5. measurement**
Length of a single page [mm					
Width of a single page [mm]					

Protocol 2

Motion with Constant Acceleration

Use a separate chronometer

Tasks

1. Estimate the initial position and the constant acceleration of a massless vehicle moving along the “x” axis.

$$x = x_0 + \frac{1}{2}at^2$$

2. Set up the theoretical problem in: https://www.walter-fendt.de/html5/phen/acceleration_en.htm with $x_{0,\mathrm{th}} = +2\,\mathrm{m}$ $\quad a_{\mathrm{th}} = +1.5\,\mathrm{m/s^2}$

3. Record N, x, and t in a table.

4. Plot distance (x) vs. time (t). Is the relation linear? If not, can you make it linear?

5. Do a linear regression using the above method to estimate the initial speed and the constant acceleration with their respective errors.

6. Are these values close to the theoretical ones?