Experiments, measurements

ver. 20200110

Models

- Representation of a selected part of the world (phenomena or data)
- Representation of a theory in the sense that it interprets the laws and axioms of that theory

These are not mutually exclusive, scientific models can be representations in both senses at the same time.

Models can be:

- physical objects
- fictional objects
- set of theoretic structures
- descriptions
- equations
- combinations of some of these

Why to use models?

Typically easier, more economical the investigation of a model than the original system.

How to create a model?

Collect data and find the contexts!

Data can be **qualitative** ones (sg likes to, sg tends to, color) or **quantitative** ones (e.g. length, weigth, temperature, wavelength, mass).

Quantitative data can be collected by **measurements**.

We can simple **observe something** or **perform an experiment**.

Experiment: a procedure we carry out to get information on the system/phenomena.

A simple observation is not reproducible. In an experiment we **control the conditions** and can reproduce the phenomena.

How to document an experiment? We write **lab reports**. In the lab reports all important information must appear.

Measurement

Measurement is the procedure which provides the measure of a physical quantity.

We can measure **directly** (e.g. the length of the rubber can be directly compared by the ruler) or **through interactions** (e.g. the volume of a liquid increases with temperature, therefore we can measure T using a liquid thermometer).

The value of the physical quantity is a product:

X = {x}[X], where X is the physical quantity, {x} is the measure, [X] is the unit.

Complex measurement systems contain a **sensor** which is the component sensitive to the quantity of interest.

Complex measurement systems can be **analog** or **digital**.

Classification of sensors

Physical sensors:

- pressure
- temperature
- movement
- light intensity
- electric conduction

Chemical sensors:

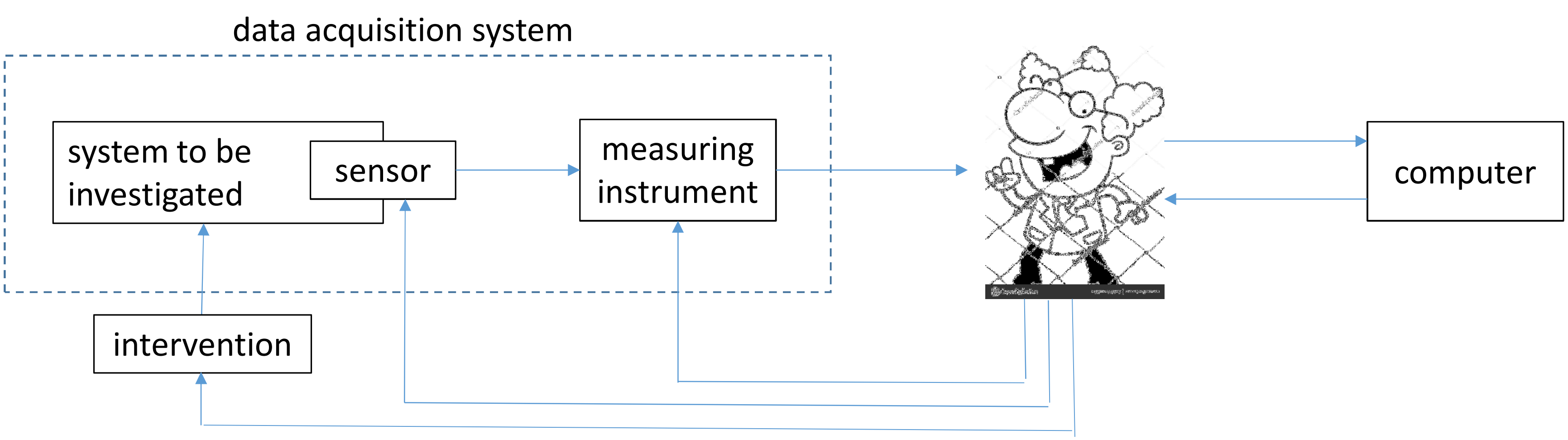
- pH
- concentration

Biological sensors (biosensors):

- antigene
- enzyme

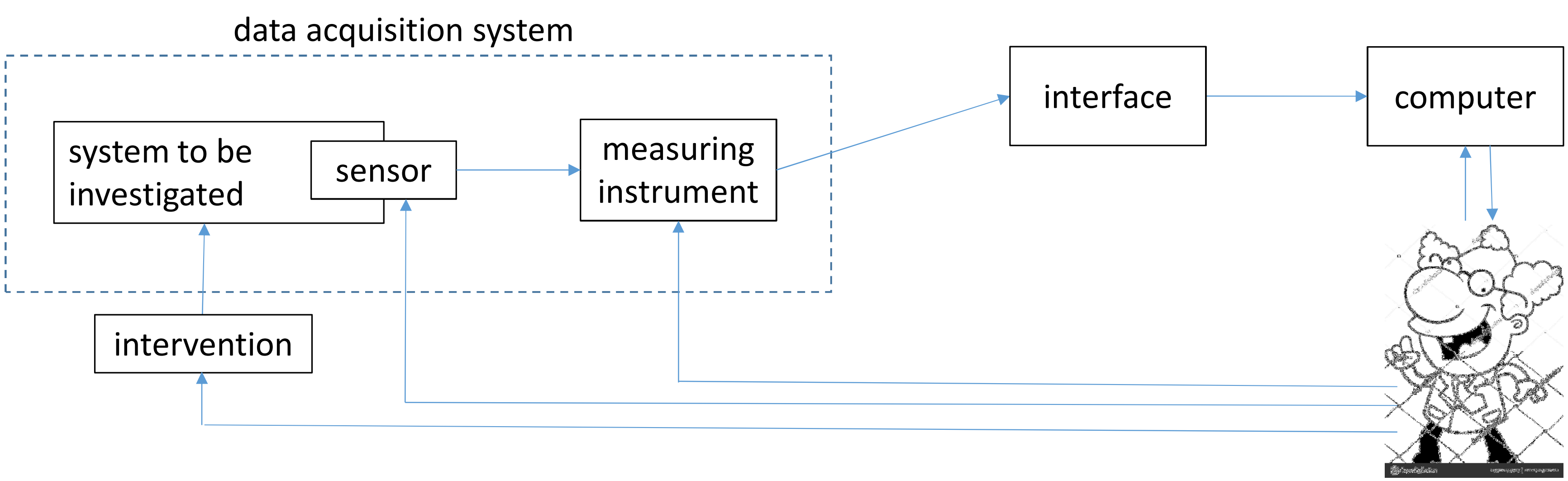
Classification of data acquisition systems and data processing

Off-line



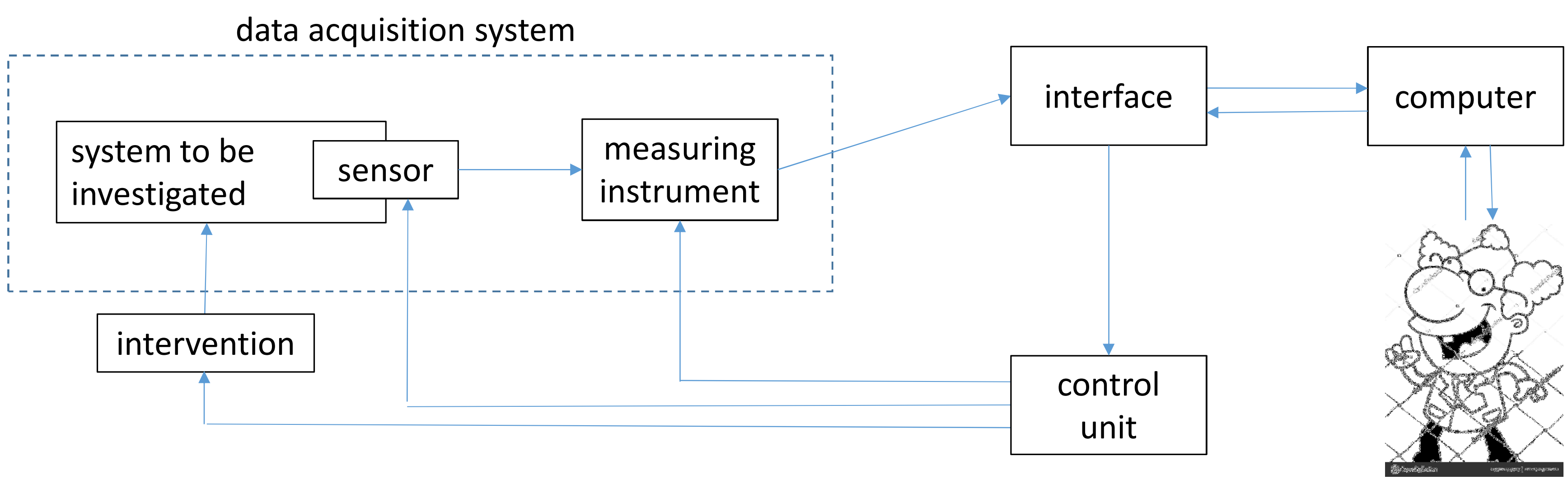
Classification of data acquisition systems and data processing

On-line



Classification of data acquisition systems and data processing

Automatic



Error of the experiments

Experimantal error (observational error, measurement error):

difference between a measured value of a quantity and its true value

Note: observational error is a misleading term, because it may refer also to a badly performed experiment.

- **Systematic error:** always occurs with the same value, when the instrument is used in the same way and in the same case. Can be constant or changing in time (drift).
  Reason: the experimental technique causes deviation from the real value
  Direction: one-way
  Elimination: impossible, can be estimated using another, different experimental method
  Related to: calibration, accuracy, bias

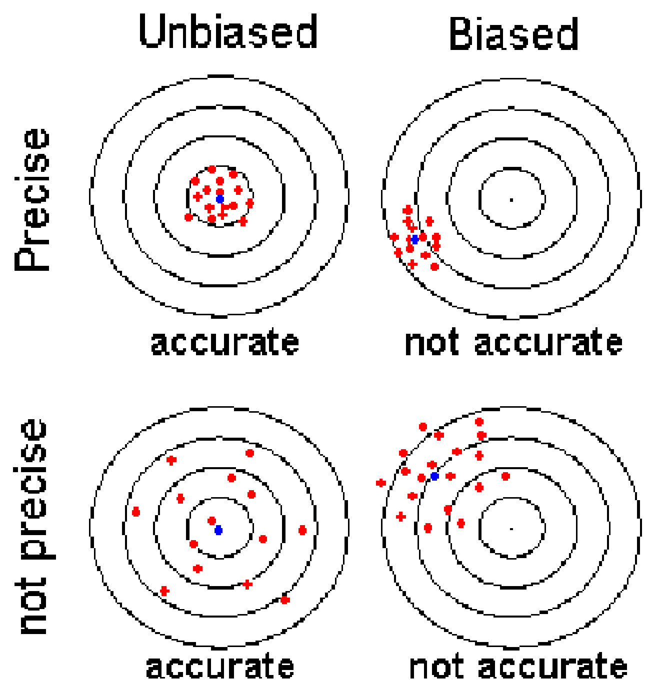
- **Random error:** may vary from observation to another
  Reason: the experimental conditions can not be controlled infinitely, therefore the measured values fluxtuate unpredictable
  Direction: variable way
  Elimination: by the repetition of the experiment
  Related to: precision

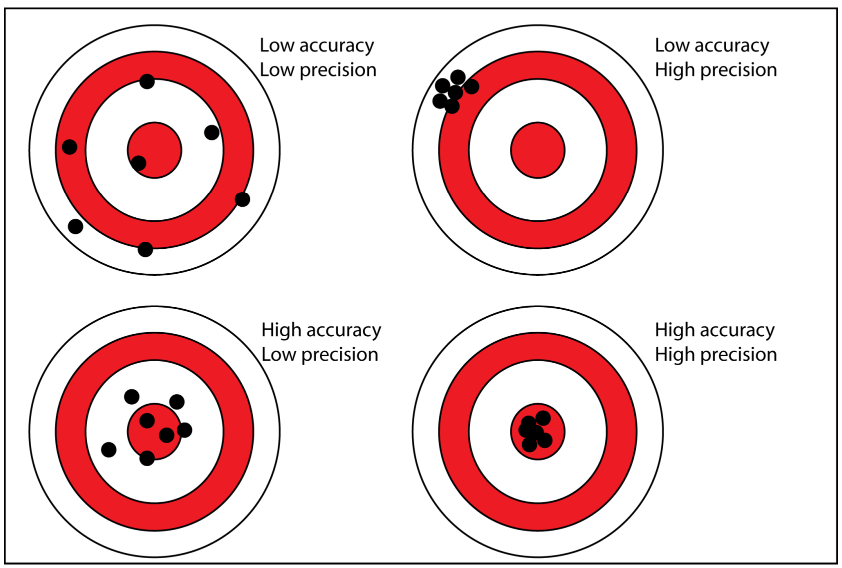
Precision and accuracy

In measurement of a set

accuracy:
refers to closeness of the measurements to the real value

precision:
refers to the closeness of the measurements to each other





Absolute error, Instument limit of error, Least count

Absolute error: $\Delta = |X - x|$
where *X* is the real value, *x* is the determined value of the quantity

Equivalently: $X = x \pm \Delta$

Dimension of Δ equals to the dimension of *X* and *x*.

Problem: we do not know the real value
Solution: we can estimate the real value or the error limit (*h*)

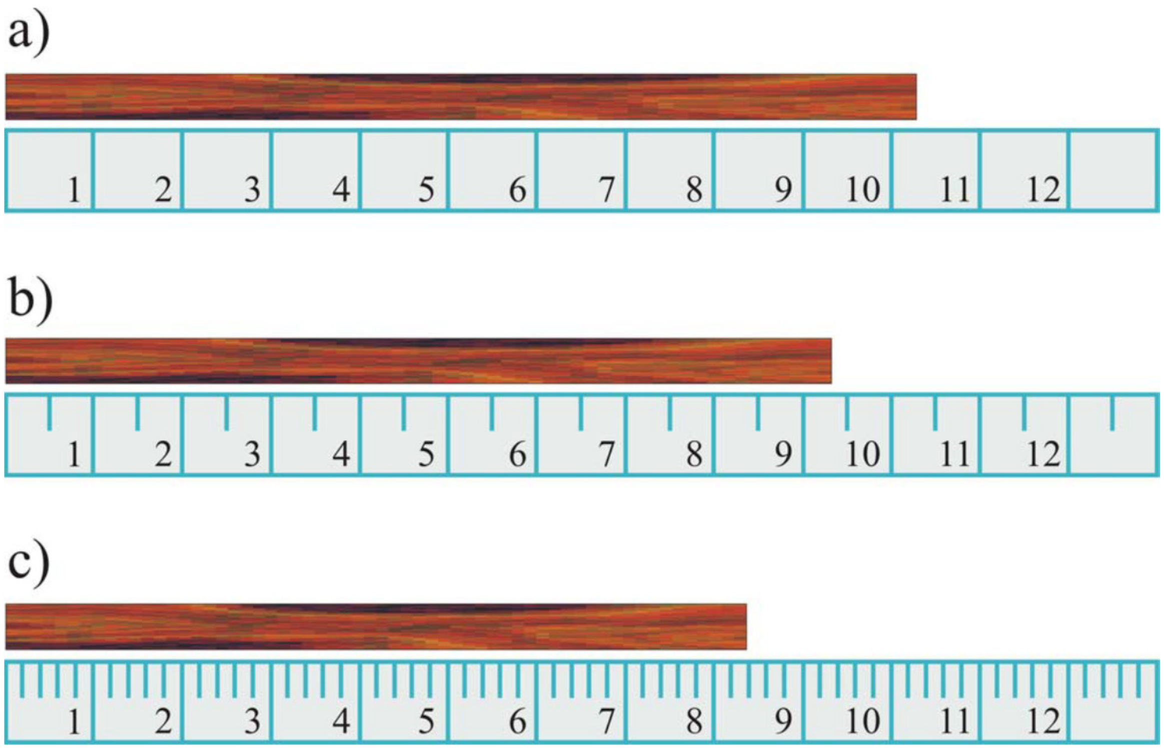
$$\Delta = |X - x| \leq h$$

Instrument limit of error: the precision to which a measuring device can be read. It is always equal to or smaller than the least count.

Least count: the resolution of the reading (in digital devices the smallest digit, in analog devices the smallest difference between two scale ticks)

Absolute error, Instument limit of error, Least count

Give the least count and the instrument limit of error for these rulers scaled in centimeters and give the lengths of the rods!



Least count	Instument limit of error	Rod length
1 cm	0.1 cm	10.3 cm
0.5 cm	0.1 cm	9.3 cm
0.2 cm	0.05 cm	8.35 cm

Relative error

Relative error: $\Delta_{\mathrm{rel}} = \frac{\Delta}{x}$

where Δ is the absolute error, *x* is the determined value of the quantity

Dimension of Δ_{rel}: dimensiononless

Problem: we do not know the absolute error

Solution: we can estimate the real value or the relative error limit (h_{rel})

$$\Delta_{\mathrm{rel}} \leq h_{\mathrm{rel}}$$

In practice: if we say error we refer to the approriate estimeted error limit.

Propagation of error

Suppose you measure some quantities *a*, *b*, *c* with error Δa, Δb, Δc.
If you want to calculate some other quantity, *Q* which depends on *a*, *b*, and *c*
what is the error of *Q*?

We feel that the error of *a*, *b*, *c* somehow propagates to *Q*.

In a simple case (*a*, *b*, *c* are independent (non-correlated)):

$$\Delta Q = \sqrt{\left(\frac{\mathrm{d}Q}{\mathrm{d}a} \cdot \Delta a\right)^2 + \left(\frac{\mathrm{d}Q}{\mathrm{d}b} \cdot \Delta b\right)^2 + \left(\frac{\mathrm{d}Q}{\mathrm{d}c} \cdot \Delta c\right)^2}$$

This is the **Gaussian Error Propagation** formula.

Sample calculation for the propagation of error

We determine the mass of solution as the sum of the solute and the solvent. Estimate the error if the mass of solute (1.5324 ± 0.0120) g and the mass of solvent (99.64 ± 0.04) g!

$$m_{\text{solution}} = m_{\text{solute}} + m_{\text{solvent}}$$

Let us apply the Gaussian Error Propagation formula:

$$\Delta m_{\text{solution}} = \sqrt{\left(\frac{\mathrm{d}m_{\text{solution}}}{\mathrm{d}m_{\text{solute}}} \cdot \Delta m_{\text{solute}}\right)^2 + \left(\frac{\mathrm{d}m_{\text{solution}}}{\mathrm{d}m_{\text{solvent}}} \cdot \Delta m_{\text{solvent}}\right)^2}$$

$\frac{\mathrm{d}m_{\text{solution}}}{\mathrm{d}m_{\text{solute}}} = 1$, $\frac{\mathrm{d}m_{\text{solution}}}{\mathrm{d}m_{\text{solvent}}} = 1$, therefore

$$\Delta m_{\text{solution}} = \sqrt{\Delta m_{\text{solute}}{}^2 + \Delta m_{\text{solvent}}{}^2} = \sqrt{0.0120\ \mathrm{g}^2 + 0.04\ \mathrm{g}^2} = 0.04176\ \mathrm{g}$$

The mass of solution: $m_{\text{solution}} = 101.1724\ \mathrm{g} \pm 0.04176\ \mathrm{g} \approx (101.172 \pm 0.042)\ \mathrm{g}$

What’s going on here?

Accuracy of a value

The accuracy of a value can be quantified by the **number of significant digits**.

Significant digits (figures): all digits except leading zeros.

Significant digits are denoted by red color:

0.001460

1.460

1460

14600

$1.460 \cdot 10^6$

$0.517 \cdot 10^{-4}$

Determination of the number of significant digits

Determine the number of significant digits in the following numbers:

$5.13470 \cdot 10^6$; $5.0 \cdot 10^{-2}$; 210; 16.40; 0.04124; 0.1463; 4140.8

Round the following numbers to the given number of significant digits:

$5.13470 \cdot 10^6$	1, 2, 3
$5.016 \cdot 10^{-2}$	3
210	2
16.40	1, 2
0.04124	2
0.1463	2
4140.8	2

Accuracy during calculations

Best way, but sometimes too difficult: perform a whole error propagation analysis and give the absolute error limit.

E.g. The mass of solution: $m_{\mathrm{solution}} = 101.172\ \mathrm{g}\ \pm 0.04176\ \mathrm{g} \approx (101.172 \pm 0.042)\ \mathrm{g}$

1. Round the **absolute error limit** to **two significant digits**.
2. Round the **result value** to **the same decimal** as the error ends.

Practical way:

1. Check the number of significant digits of the given data.
2. For intermediate numbers use two more significant digits than the given data had.
3. For the results give as many significant digits as that of the given data (the smallest one).

On the reliability of experiments

The reliability of the experiments from the quantitative point of view is investigated by the **mathematical statistics**.

All possible results of an experiment is called **statistical population**. A subset of the statistical population is a **statistical sample** (a set of experiments).

Nobody can know the whole population, but **can investigate samples on**ly and draw conclusions from the features of a sample to the whole population.

A random variable (or stochastic variable) is a variable whose values depend on outcomes of a random phenomenon. The result of an experiment is a random variable.

Probabilities

If we have a continous random variable one of the most important questions is the probability of being its value in a given interval.

This probability is described by the **probability density function**.

Typical probability density functions (distributions):

- normal (or Gaussian)
- Poisson
- even

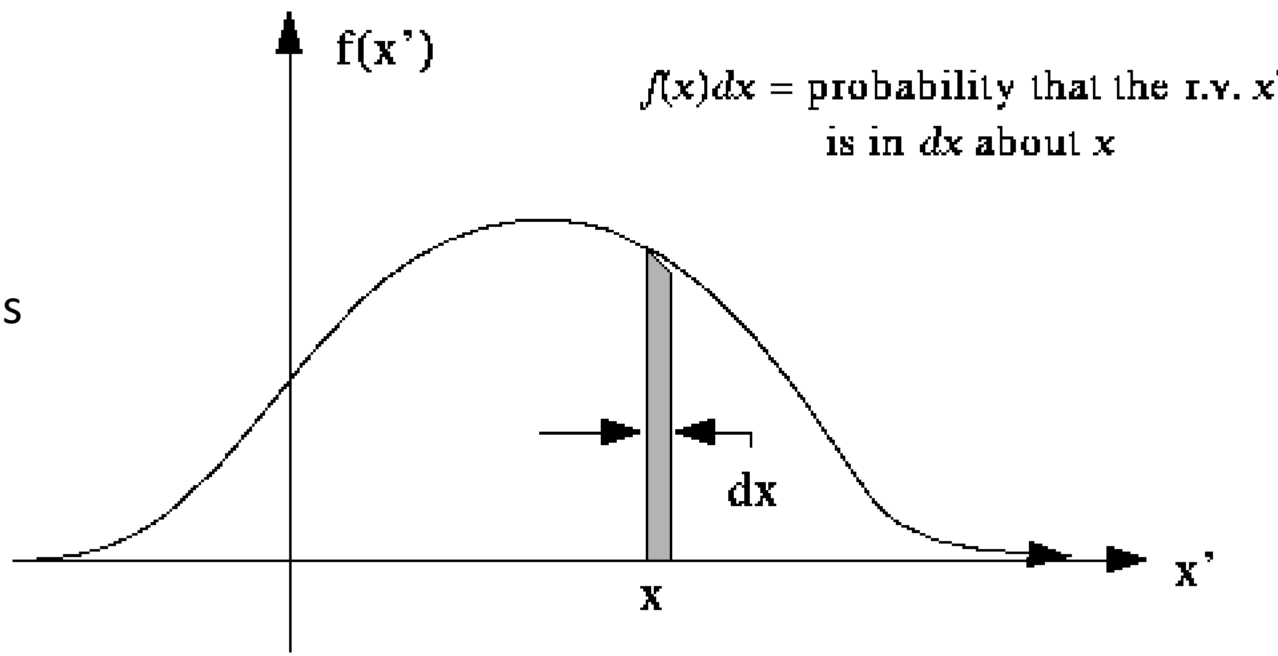


Figure 4. Typical Probability Distribution Function (*pdf*)

The normal distribution

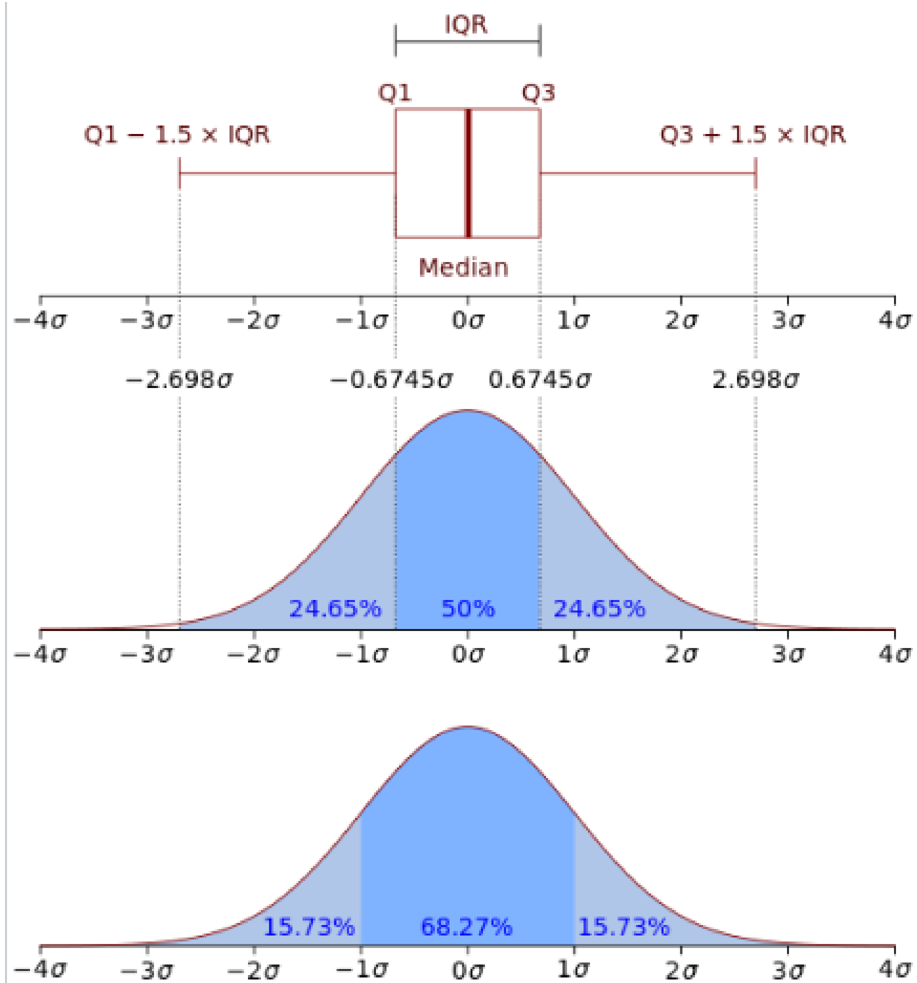
Normal distribution is quite common in the nature when several source of errors join and each has small effect on the measured value.

The normal distribution can be described by two parameters:

μ is the mean or expectation of the distribution

σ is the standard deviation (σ>0)

σ^2 is called variance



Estimation of the paramters of the normal distribution

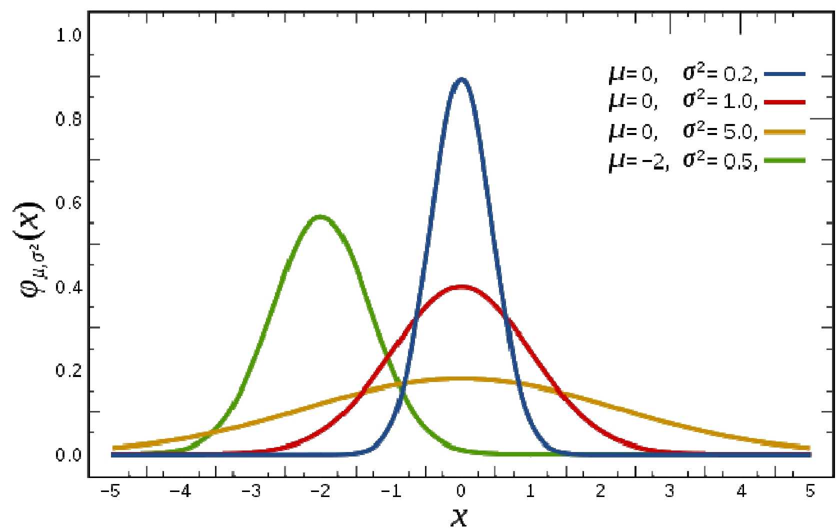
We do not know the **real value** of a quantity and if we assume its distribution to be normal the mean could be a good estimation of the unknown real value.

The mean of a normal distribution can be estimated by the **arithmetic mean**: $\bar{x} = \frac{\sum_{\mathrm{i}=1}^{n} x_{\mathrm{i}}}{n}$

The **standard deviation** gives information about the **reproducibility** of the experiments and the **average deviation from the mean**.

Smaller standard deviation: more reliable experimental results.

We do not know the standard deviation, but can be estimated by the **corrected sample standard deviation with Bessel's correction**: $\sigma = \sqrt{\frac{\sum_{\mathrm{i}=1}^{n}(x_{\mathrm{i}}-\bar{x})^2}{n-1}}$



The confidence interval

The **confidence interval** is a statistically estimated interval that contain the true value of an unknown population parameter with a given level of confidence.

Estimation of the confidence interval: $k \cdot \sigma$
where σ is the standard deviation and k is a multiplicator factor depending on the level of confidence (**for large number of experiments** (>30)).

For less number of experiments the Student's t distribution should be used and our formula is very similar: $t_\alpha \cdot SE$
where SE is the **standard error** and t_α is a multiplicator factor (critical value) depending on the **level of confidence** and the **degrees of freedom**.

Degrees of freedom (*f*): f = number of datapoints – number of determined parameters.

How to give a result with confidence interval

Steps of work for small number of experiments:

- Determine the number of experiments, the mean, the standard error and the degrees of freedom.
- Select a level of confidence and find the appropriate critical value correspoding the degrees of freedom.
- Calculate the confidence interval and round it to two significant digits.
- Round the mean value to the same decimal where the confidence interval ends.

How to give a result with confidence interval

a case study

We measured the temperature of an eutectic system 12 times.

The mean is estimated as the arithmetic average: $\bar{T} = 180.911\ °\mathrm{C}$, the calculated standard deviation $\sigma = 1.6686\ °\mathrm{C}$

The standard error of the average is calculated dividing the standard deviation by the square root of the number of datapoints: $SE = \frac{\sigma}{\sqrt{n}} = \frac{1.6686\ °\mathrm{C}}{\sqrt{12}} = 0.481683\ °\mathrm{C}$

How to give a result with confidence interval
a case study

The degrees of freedom, $f = n - 1 = 11$ (we have one estimated parameter, the mean).

We select 95% confidence level.

Let us determine value of critical level!

95% confidence level -> 5% significance level in this table.

Therefore $t_\alpha = 2.201$

Degrees of freedom	Significance level					
	20% (0.20)	10% (0.10)	5% (0.05)	2% (0.02)	1% (0.01)	0.1% (0.001)
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.598
3	1.638	2.353	3.182	4.541	5.841	12.941
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.859
6	1.440	1.943	2.447	3.143	3.707	5.959
7	1.415	1.895	2.365	2.998	3.499	5.405
8	1.397	1.860	2.306	2.896	3.355	5.041
9	1.383	1.833	2.262	2.821	3.250	4.781
10	1.372	1.812	2.228	2.764	3.169	4.587
11	1.363	1.796	2.201	2.718	3.106	4.437
12	1.356	1.782	2.179	2.681	3.055	4.318
13	1.350	1.771	2.160	2.650	3.012	4.221
14	1.345	1.761	2.145	2.624	2.977	4.140
15	1.341	1.753	2.131	2.602	2.947	4.073
16	1.337	1.746	2.120	2.583	2.921	4.015
17	1.333	1.740	2.110	2.567	2.898	3.965
18	1.330	1.734	2.101	2.552	2.878	3.922
19	1.328	1.729	2.093	2.539	2.861	3.883
20	1.325	1.725	2.086	2.528	2.845	3.850

How to give a result with confidence interval

a case study

We measured the temperature of an eutectic system 12 times.

The mean is estimated as the arithmetic average: $\bar{T} = 180.911\ °\mathrm{C}$, the calculated standard deviation $\sigma = 1.6686\ °\mathrm{C}$

The standard error of the average is calculated dividing the standard deviation by the square root of the number of datapoints: $SE = \frac{\sigma}{\sqrt{n}} = \frac{1.6686\ °\mathrm{C}}{\sqrt{12}} = 0.481683\ °\mathrm{C}$

The degrees of freedom, $f = n - 1 = 11$ (we have one estimated parameter, the mean).

We select 95% confidence level, therefore the critical level, $t_\alpha = 2.201$

The confidence interval: $t_\alpha \cdot SE = 2.201 \cdot 0.481683\ °\mathrm{C} = 1.06019\ °\mathrm{C} \approx 1.1\ °\mathrm{C}$

$$T = \bar{T} \pm t_\alpha \cdot SE = (180.9 \pm 1.1)\ °\mathrm{C}$$