Introduction to error analysis

Advanced Methods in Bioengineering Laboratory

No measurement made is ever exact!

- The **accuracy** (correctness) and **precision** (number of significant figures) of a measurement are always limited by the degree of refinement of the apparatus used, by the skill of the observer, and by the basic physics in the experiment.
- In doing experiments we are trying to **establish the best values** for certain quantities, or trying to validate a theory.
- We must also give **a range of possible true values** based on our limited number of measurements.

What you need to know for this lab:

- Different error types and sources
- How to calculate a standard deviation
- How to treat outliers
- When to use standard error and when standard deviation
- Simple error propagation

Different types of errors

- **Systematic errors:** Errors that occur due to a fundamental problem in the measurement (for example instrument calibrated wrong). *Solution: improve experimental design and use systematic correction.*

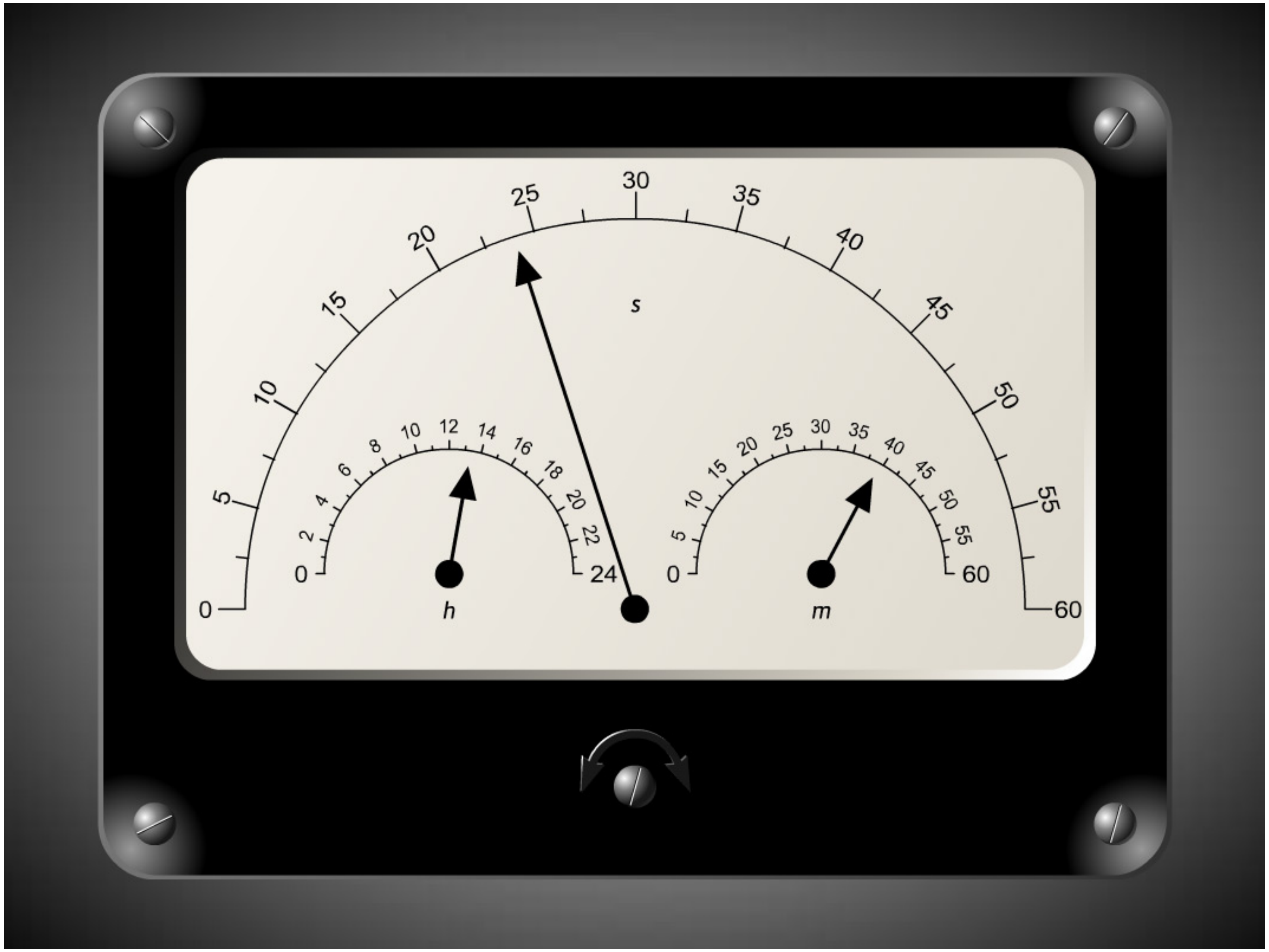
- **Random errors:** Errors that occur randomly during the measurement and result in a statistically non correlated error for each time the measurement is made. *Solution: statistics!*

Determining random errors

The instrument you are using contributes to the random error. We can describe how much it contributes with two quantities:

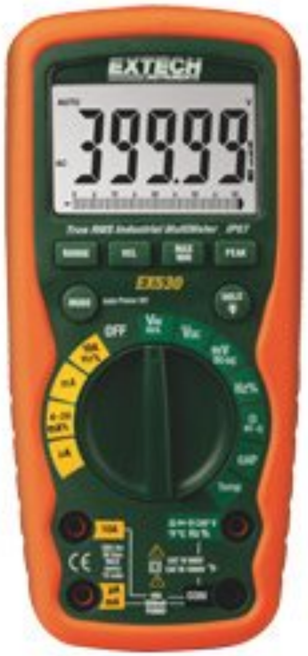
- **Instrument Limit of Error (ILE):** the precision to which a measuring device can be read, and is always equal to or smaller than the least count.
- **Least Count:** the smallest division or digit marked on an instrument

Example: Analog Instruments



- Least count= 2.5s
- Instrument limit of error≈ 1s

Example: Digital Instruments



- Instrument limit of error (ILE)= Least count
- BEWARE: Not all digits are full digits
- BEWARE: Not all digits are meaningful

Estimated Uncertainty by Repeated Measurements

When you've done a measurement, you want to know how close your measurement most likely is to the "real" value.

Solution: Repeat the measurement several times, find the average, and find either the average deviation or the standard deviation.

Average		Average deviation	Standard deviation
Time t/s	$(t-<t>)/s$	$\vert(t-<t>)\vert/s$	$(t-<t>)^2/s^2$
7.4	-0.2	0.2	0.04
8.1	0.5	0.5	0.25
7.9	0.3	0.3	0.09
7.0	-0.6	0.6	0.36
$<t>=$ 7.6	$<t-<t>>=$ 0	$<\vert t-<t>\vert>=$ 0.4	$\left\langle (t-<t>)^2 \right\rangle =$ 0.247
		standard deviation	$\sqrt{\left\langle (t-<t>)^2 \right\rangle} =$ 0.5

How to calculate the standard deviation?

1. Compute the square of the difference between each value and the sample mean.
2. Add those values up.
3. Divide the sum by n-1. This is called the variance.
4. Take the square root to obtain the Standard Deviation.

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

x. . . each score
$\bar{x}$. . . the mean or average
n. . . number of values
$\sum$. . . means the sum over the values

Rules for treating outliers:

- Use critical judgment!
- Think if there could be a reason that the outlier is a valid data point
- Rule of thumb: if a value is removed more than 4 times the standard deviation from the mean (calculated from the other values, disregarding the potential outlier), then you might disregard the data point in your further analysis.

Why make many measurements? Standard Error in the Mean.

Often we make many measurements and the average the measurements to reduce the "noise" (= the random errors). There are two distinctly different cases when we do this. Those cases also decide if we report our uncertainty as standard deviation or as standard error:

- The standard deviation (SD) is how spread out THINGS in the population are, and this is calculated (somehow) from the data in your sample. It is useful in describing the population itself.

- The standard error (SE) is how spread out the SAMPLE MEAN will be around the true population mean. It is useful in describing how close your results will be to the right answer.

When to use SD and when SE

- Decide if we are measuring one value multiple times (use standard error), or if we are measuring one quantity in multiple cases (use standard deviation).

- Another way to decide if you imagine you could make a perfect measurement, would you always get the same number, then use the standard error.

Examples

- When we describe a population of cells, by measuring their length, we will calculate the mean and the standard deviation, because there is no ”right length”.

- When we want to measure the temperature in our incubator, we will calculate the average temperature and the standard error, because there is a ”right temperature”.

How are Standard Error and Standard Deviation Related?

- The standard error in the mean in the simplest case is defined as the standard deviation divided by the square root of the number of measurements.

$$SE = \frac{SD}{\sqrt{n}}$$

Relative vs. Absolute Errors

When we know the error of our measurement, we can report it in different ways:

- absolute value of the error (absolute error), for example $R = (33 \pm 1.65)\mathrm{k}\Omega$
- Relative error: absolute error/value
- error percentage $R = (33 \pm 5\%)\mathrm{k}\Omega$

These three types of errors are often used in different fields. Absolute errors are often used to depict measurement errors, whereas percentage errors are often used for tolerances of parts or components. The relative error is useful for error propagation.

Propagation of Errors

Often we have to apply math to our measurement values to calculate the quantity we really want to know. Since our measurement values have errors, we know that the calculated quantities will have errors. How do we estimate the error of the calculated quantity?

- Given $x=(x\pm\Delta x)$ and $y=(y\pm\Delta y)$, what is $z(x,y)$

There are two ways to get an estimate for the error of z.

- simplified version: the guiding principle in all cases is to consider the most pessimistic situation.

- proper statistical treatment of error propagation: use the standard deviations to calculate the resulting uncertainty

Addition and Subtraction

Using average errors	Using standard deviations
$\Delta z = \lvert\Delta x\rvert + \lvert\Delta y\rvert + \ldots$	$\Delta z = \sqrt{(\Delta x)^2 + (\Delta y)^2 + \ldots}$

- Example: $w = (4.52\pm0.02)$cm, $x = (2.0\pm0.2)$cm, $y = (3.0\pm0.6)$cm. Find $z = x+y-w$ and its uncertainty.
- $z = x + y - w = 2.0 + 3.0 - 4.5 = 0.5$cm
- For the simplified method we get: $\Delta z = \Delta x + \Delta y + \Delta w = 0.2 + 0.6 + 0.02 = 0.82$ rounding to 0.8 cm:

  So $z = (0.5 \pm 0.8)$cm

- When using the standard deviation we get: $\Delta z = \sqrt{(0.22^2 + 0.62^2 + 0.022^2)} = 0.633$:

  $z = (0.5 \pm 0.6)$cm.

Multiplication by an exact number

- When multiplying a measurement value with an exact number, multiply the uncertainty also with the exact number.
- Example: The radius of a circle is $r = (3.0 \pm 0.2)$cm. Find the circumference and its uncertainty

$$C=2\pi r = 18.850\text{cm}$$

$$\Delta C=2\pi\Delta r = 1.257\text{cm}$$ (The factors of 2 and π are exact)

$$C=(18.8 \pm 1.3)\text{cm}$$

Multiplication and Division:

Using average errors	Using standard deviations
$\frac{\Delta z}{z} = \frac{\Delta x}{x} + \frac{\Delta y}{y} + \dots$	$\frac{\Delta z}{z} = \sqrt{\left(\frac{\Delta x}{x}\right)^2 + \left(\frac{\Delta y}{y}\right)^2 + \dots}$

Example: $w = (4.52 \pm 0.02)cm, x = (2.0 \pm 0.2)cm$. Find $z = w \cdot x$ and its uncertainty.

Using the standard deviations:

$$
\begin{aligned}
\frac{\Delta z}{9.04cm^2} &= \sqrt{\left(\frac{0.02cm}{4.52cm}\right)^2 + \left(\frac{0.2cm}{2.0cm}\right)^2} = 0.1 \\
\Delta z &= 0.9cm^2 \\
\text{therefore} & \\
z &= (9.0 \pm 0.9)cm^2
\end{aligned}
$$

Products of Powers $z = x^m + y^n$

Using average errors	Using standard deviations
$\frac{\Delta z}{z} = \lvert m \rvert \frac{\Delta x}{x} + \lvert n \rvert \frac{\Delta y}{y} + \ldots$	$\frac{\Delta z}{z} = \sqrt{\left(\frac{m\Delta x}{x}\right)^2 + \left(\frac{n\Delta y}{y}\right)^2 + \ldots}$

Mixtures of multiplication, division, addition, subtraction, and power

- Treat the math of the uncertainties in the same order as you would treat the math itself!

Example: $w = (4.520.02)cm, x = (2.00.2)cm, y = (3.00.6)cm$. Find $z = wx + y^2$

First we compute $v = wx$ to get $v = (9.0 \pm 0.9)cm^2$
Next we compute $\Delta(y^2)$:

$$\begin{aligned}
\frac{\Delta(y^2)}{y^2} &= \frac{2\Delta y}{y} = \frac{2 \cdot 0.6cm}{3.0cm} = 0.40 \\
\Delta(y^2) &= 0.40(9.00cm^2) = 3.6cm^2
\end{aligned}$$

finally we compute

$$\begin{aligned}
\Delta z &= \Delta v + \Delta(y^2) = 0.9 + 3.6 = 4.5cm^2 \Rightarrow 4cm^2 \\
z &= (18 \pm 4)cm^2
\end{aligned}$$

Significant Digits

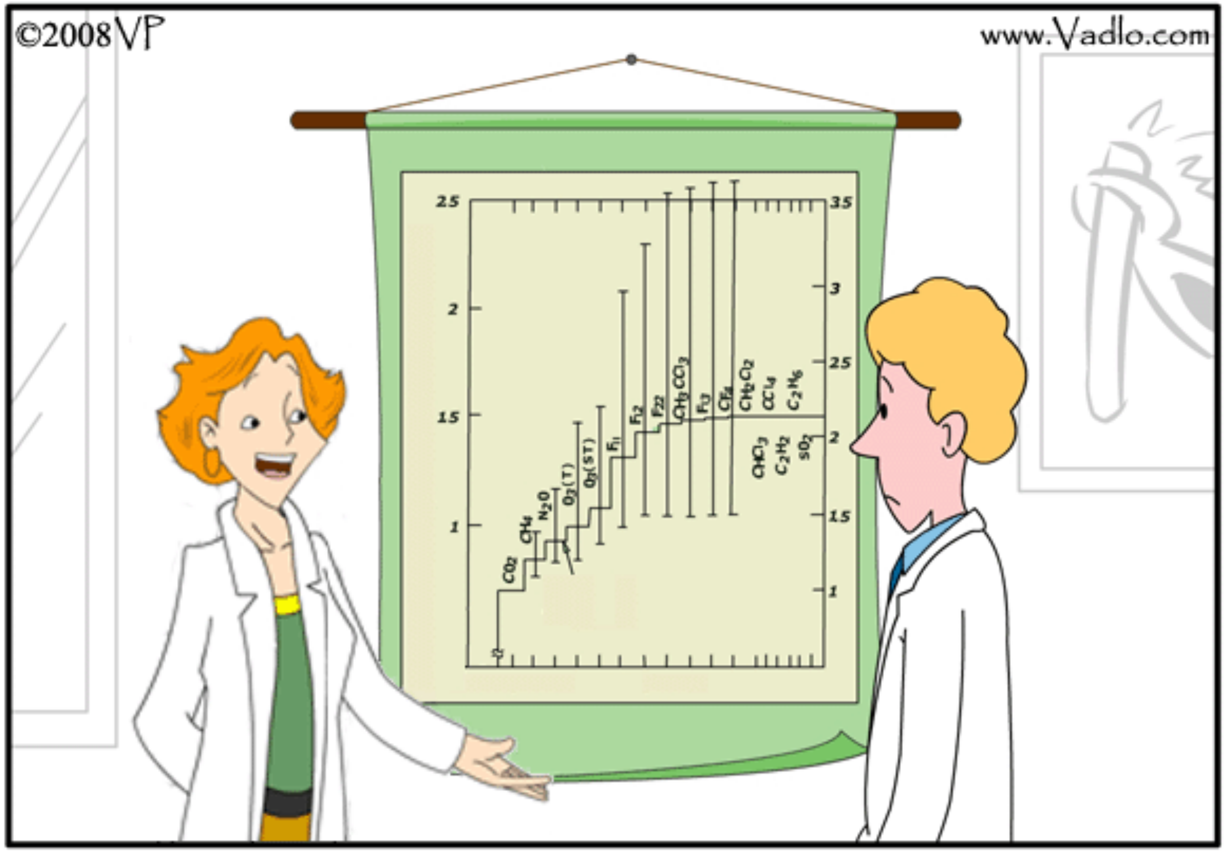
After our measurements or calculations, we often end up with a number with many digits. How many digits should we report? Which digits still have a significant meaning?

- A significant figure is any digit 1 to 9 and any zero which is not a place holder.
- Examples:
  - 1.350 there are 4 significant figures
  - 0.00320 there are 3 significant figures
  - 1350 there are ??? Significant figure
- use scientific notation: 1.35×10^{3} has 3 significant digits

How many significant figures should be in the final answer?

- In doing running computations we maintain numbers to many figures, but we must report the answer ONLY to the proper number of significant figures.
- The short rule for multiplication and division is that the answer will contain a number of significant figures equal to the number of significant figures in the entering number having the least number of significant figures.
- For measured values we report the standard deviation or standard error with one significant digit, unless the first significant digit is a 1. Then we report 2 significant digits. The measured value is reported with the same significant digits as the standard deviation or standard error.

An answer without error estimates is meaningless (and therefore WRONG)!



Did you really have to show the error bars?