Academic research and writing skills

Steve Quarrie

Consultative Bureau for International Projects - Ministry of Science and Technological Development

I have been doing research since 1967, when I started my PhD.

I have around 2000 citations of my publications so far.

I trained in organic chemistry. I have done research in: organic chemistry, biochemistry, plant physiology, quantitative genetics, hydrology, agro-economics, rural development, education.

I used to work at the John Innes Centre, UK, but now I am retired and live in Belgrade (**Brain Gain**)!

I have been teaching this type of training course since 2003, but every course is different!

I am a professor at the Universities of Belgrade and Newcastle. I also used to teach at Universities of East Anglia (UK) and Parma (Italy). I still do research at Newcastle University and around Europe.

This course has something for every researcher!
Over 1000 Serbian scientists have been on my training courses so far.

You need to appreciate what is meant by good quality to carry out good quality research!
But, to make a successful career in research you need to do a lot more than just carry out experiments!

This 1.5-day course will give you a broad range of generic (general, lifelong) skills, valuable for any career in academia and industry.

You will sometimes find extra slides in the handouts.

Stop me whenever you like to ask questions, discussion, etc.

The objective of this course:

The objective of the course is to provide you with an appreciation of the skills needed to develop a successful career in your particular areas of research.

That means primarily knowing how to do good quality research. **Without this you will never be able to succeed in publishing your work**.

Learning outcomes of the course:

***Today, 31st January* - Good quality research:**
Understanding what the objective of research should be
Understanding the importance of considering all the factors
Understanding principles of good experimental design
Understanding the importance of thinking before doing
Learning how to reduce sources of error
Appreciating the importance of data quality control.

Do's and don'ts in the data analysis process - from Ivo van der Lans

Scientific writing:
Recognising when you have good quality science
Understanding that papers must have international interest
Understanding the requirements for different types of papers

Tuesday, 1st February, morning **- Scientific writing continued:**
Learning how to improve the text
Learning how to improve your English
Learning how to prepare an abstract

Writing proposals:
Understanding the philosophy for success
Learning how to be competitive
Learning how to read the instructions
Understanding how to give enough details
Understanding how to be consistent in what you say
Understanding what is needed for effective project management
Learning how to prepare an effective budget
Learning how to present your proposal effectively

Research philosophy**:**

Throughout this course I shall be describing series of thought processes - sequences of events.

The secret to being a successful researcher is
- **knowing how to think logically:**
- **never believing anything on principle!**

- **knowing how to think logically:**
  - being able to put your thoughts down on paper **logically**,
  - being able to develop an idea **logically**,
  - being able to interpret your data **logically**,
  - being able to communicate your work **logically** to others (through written and oral forms).

- **never believing anything on principle:**
  - don't believe anything anyone tells you,
  - don't believe any of your experimental data,
  - the "default" setting should be **don't believe anything unless it is proved to your satisfaction!**

“Clearly, scientists whose work is never cited should seriously consider doing something different with their lives.”

A quotation from: Federico Rosej and Tudor Johnston
Survival Skills for Scientists (2006)
Imperial College Press [ISBN 1-86094-641-0 (pbk)]

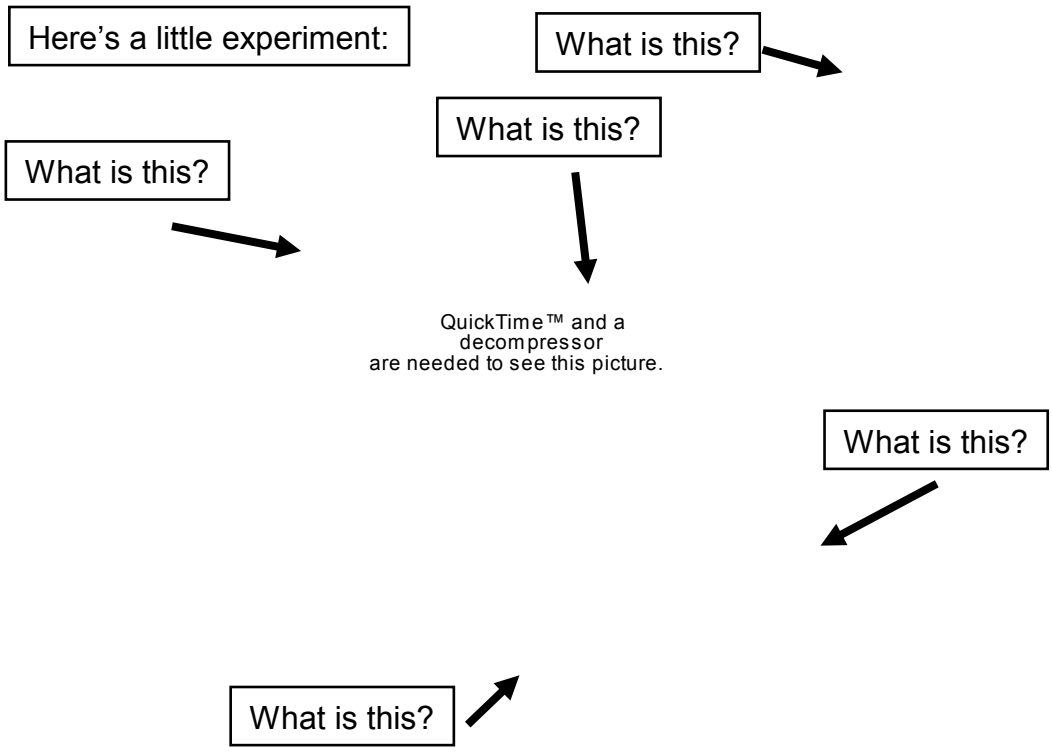
Ask yourself the questions -
“Will other scientists be interested in my research?”
“Will my research have any impact on anything or anyone?”

If your answers to both questions are “No”, then you need to apply what you learn on this course and change your approach to doing research,

or consider doing something different with your life!

A further reading list:

1. Survival Skills for Scientists (2006)
   F. Rosej, T Johnston. Imperial College Press, ISBN 1-86094-641-0 (pbk)
2. Experimental Design for the Life Sciences (2006)
   G.D. Ruxton, N Colegrave, Oxford Univ. Press, ISBN 0-19-928511-2
3. How to Write & Publish a Scientific Paper (1998) 5th Edition
   R. A. DayOryx Press, ISBN 1-57356-165-7 (pbk)
4. How to Write and Illustrate a Scientific Paper (2003)
   Björn Gustavii, Cambridge University Press, ISBN 0-521-53024-5 (pbk)
5. Secrets to success with FP7 REGPOT proposals (2011)
   Steve Quarrie, steve.quarrie@nauka.gov.rs
6. How to Write a Competitive Proposal for Framework 7 (2006)
   Sean McCarthy, Watermans Printers Ltd. ISBN 0-9546257-2-2



Be careful when you state what the truth is!

As scientists doing research, there may frequently be no absolute truth where you can make the statement
"This is the truth".

Instead, you should use your judgement to decide when to say
"**This is the truth**", or whether you should say
"This looks like the truth" or
"I believe this to be the truth" or
"I think I've got as close to the truth as I can get."

Or perhaps:
"I forgot to consider that, so maybe it's not the truth!"

Here's how a court of law would find the truth:

The case of the Canadian Cheese Burglar!

Int J Legal Med (1999) 112: 201–203

Case circumstances:

Two suspects were caught within 12 h of a robbery and were found to be in possession of articles taken from a private residence.

The suspects claimed they were innocent of any crime.

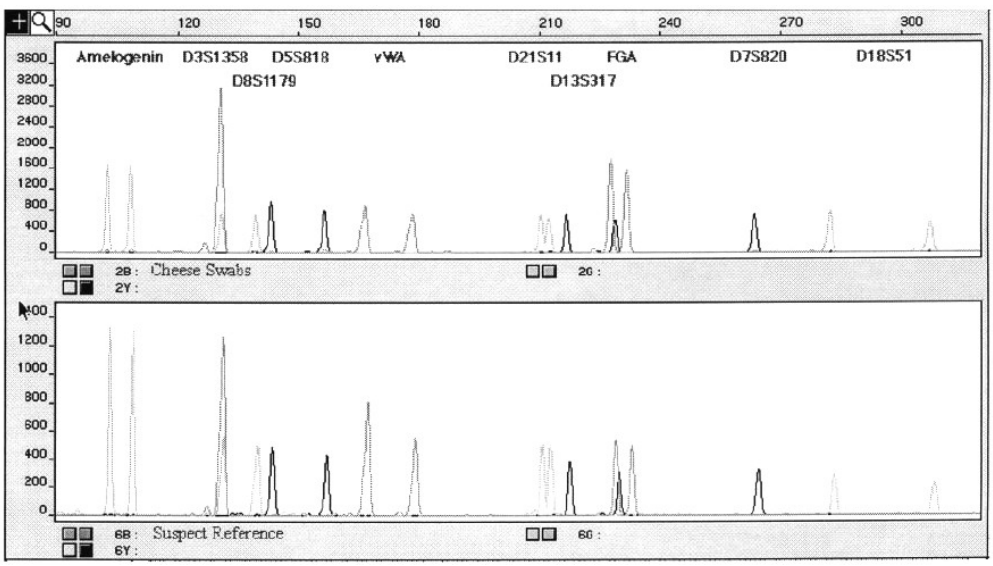
Although evidence showed that the suspects possessed stolen property, there was no evidence connecting the suspects to the crime scene.

With the help of the home owners, police recovered a small piece of cheddar cheese from the scene within 36 h which showed bite marks, and human DNA was recovered from saliva on the bite marks.

Here is the cheese at the centre of this case:

Police got blood samples from the two suspects, and DNA was extracted from the blood.

Here are electropherograms comparing DNA recovered from saliva on the cheese (upper traces) with DNA from one of the suspects (lower traces) using markers for 10 human DNA sequences.



DNA collected from the cheese and suspect was identical at every position.

Markers for 10 DNA loci (chromosome locations) gave identical results for DNA taken from one of the suspects and from the cheese bite marks.

Calculating the frequency of this genotype in the Canadian population gave 1 chance in 1.59 x $\mathbf{10^{14}}$.

That is 1 person out of 159000 **billion** people.
With a world population for January 2011 estimated to be 7.1 billion, that is about 22000 times the entire population of planet Earth!

That is a highly significant result!

So, the DNA from both the surface of the cheese and the suspect was identical.

So, here was the truth.

So, just like the court case of the Canadian cheese burglar, you have a lot of responsibility to get it right!

- *to identify the truth* about what really happened.

However, the truth may not always be easy to find, and the truth may not always be what you think it is!

When you do research, your search for the truth must take ***first priority***.

If you don't, you will have trouble publishing your work in peer-reviewed international journals!

So, just like a **court of law**, you have a lot of responsibility to get it right when doing research!
If you don't, your research career will suffer!
You should compare yourself with all the participants of a court trial.
When doing research you are:
- ♦ Counsel for the prosecution
- ♦ Counsel for the defence
- ♦ All the expert witnesses
- ♦ The judge and (in the UK), also
- ♦ The jury

- all at the same time!

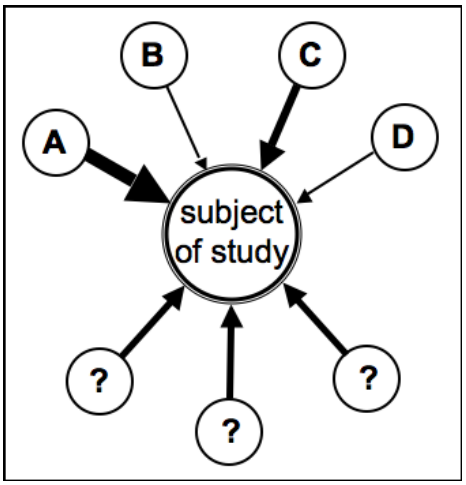
So, how do you get to the truth of your research?
By considering all the "factors".

> **factor** |ˈfaktər|
> noun
> **1** a circumstance, fact, or **influence that contributes to a result or outcome** *: his legal problems were not a **factor in** his decision | she worked fast, conscious of the time factor.*

Thus, the more you can take account of all the "factors", the closer you get to the truth. So, the more you can take account of all the factors, the better the *quality* of your research.

You can get to the ***truth*** of your research ***only*** if you can identify, and establish the relative importance of, ***all*** the influences on the experimental subject: all of the ***factors***.

Conceptually, research will generally be carried out in one of two forms.

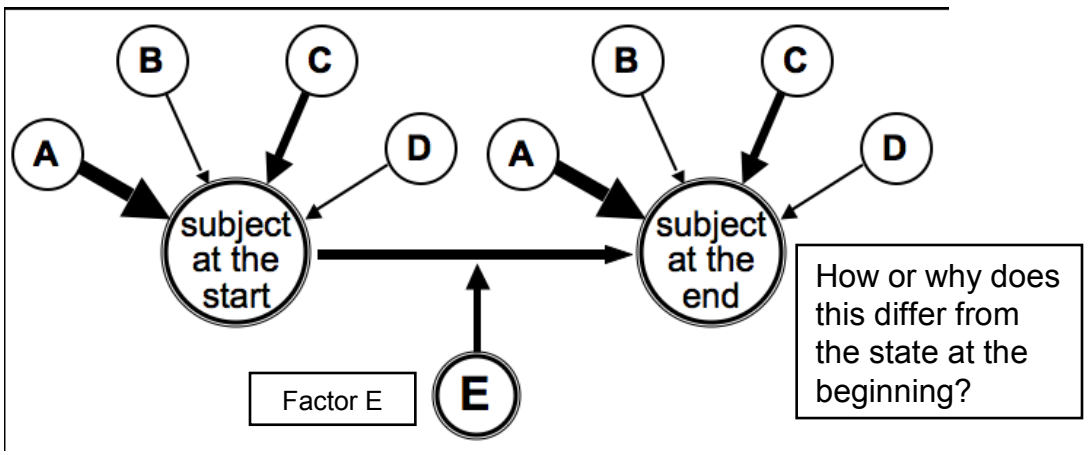


Here is the first form:

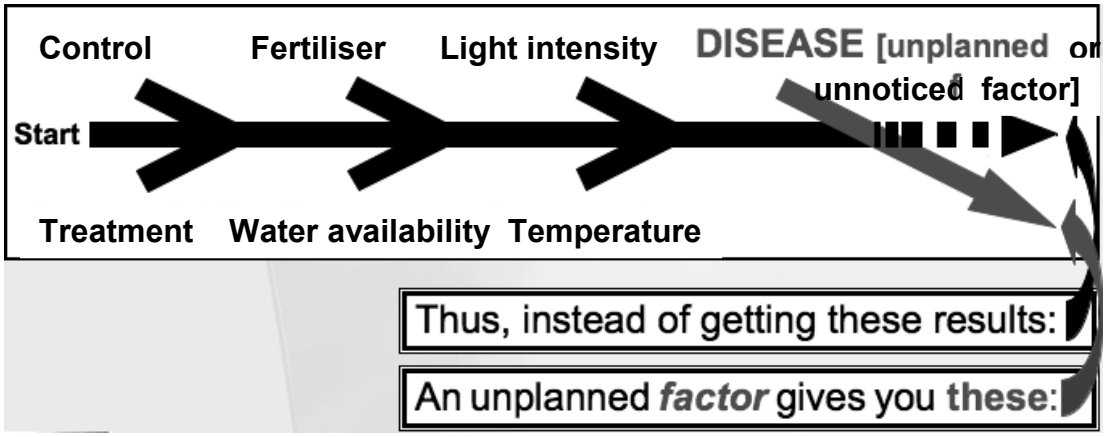
The objective of the research is to identify and quantify all the ***factors***.

The second type of research is studying the consequence of some defined change in the subject of study.

Here the objective of the research is to determine the consequence(s) of a change in these defined ***factors***.



Here is an example of the second type of research illustrated schematically for an experiment on plants, to study the consequence of changing one ***factor*** (treatment):



Note that this principle (philosophy) will be the same whatever the subject of the research: biology, economics, psychology, etc …

So, if you don't think carefully enough about all the ***factors*** that ***might*** have an influence on your research (subject of study), then you will not get to the ***truth*** of what happened/what is happening.

Good research is based on identifying (and quantifying) all these ***factors*** to get to the ***truth***.

You identify these factors in various ways:
You were taught them as an undergraduate.
You found out about them from the literature.
Your supervisor told you about them.
You gradually got the skill to recognise them.
You think there could be others you need to test.

You test these factors by setting up ***hypotheses.***

Thus, good quality research is based on answering questions in the form of ***hypothesis*** testing..

A dictionary definition

Hypothesis [n]
1. A concept that is not yet verified but that if true would explain certain facts or phenomena;
2. A proposal intended to explain certain facts or observations

Here's how it would work -

You start by asking a question. For example: Why don't official statistics on smokers in Serbia agree with my own observations?

You set up a ***hypothesis*** and then design your research/an experiment to test this:

Hypothesis: statistics are unreliable because questionnaires are badly designed.

Testing your ***hypothesis*** then becomes the objective of your research/experiment.

The quality of your research will determine how effective you are at testing your ***hypotheses***.

If your hypothesis requires you to test the consequence of varying a particular factor (for example, heavy metal toxicity in the plant experiment) ***keep other factors constant.***

Changing more than one factor at a time will invalidate your test of your hypothesis.
Testing your hypothesis under varying (natural) conditions - where you cannot control all the factors - is more of a challenge!

Extra information:
Decreased water availability causes ABA* synthesis and this increase in ABA modifies gene expression via an ABA-mediated signalling pathway.

* ABA is a plant hormone that is produced when plants are water stressed.

The hypothesis implies that stress increases ABA concentrations, and this ABA change **then** modifies gene expression.

The hypothesis tells you what you need to look at (***factors* to quantify**):
- varying the availability of water
- measuring ABA concentrations
- aspects of gene expression

The project will look at varying the availability of water as a *treatment*,
[- suitable amounts of water for plants will need to be considered.]
then look at the consequences of this in terms of ABA production.
[- decisions will need to be made on where to measure ABA and when.]
Finally, some aspects of gene expression will need to be studied.
[- decisions will need to be made on which genes and when to measure gene expression.]

However, this last point raises a critical additional question:

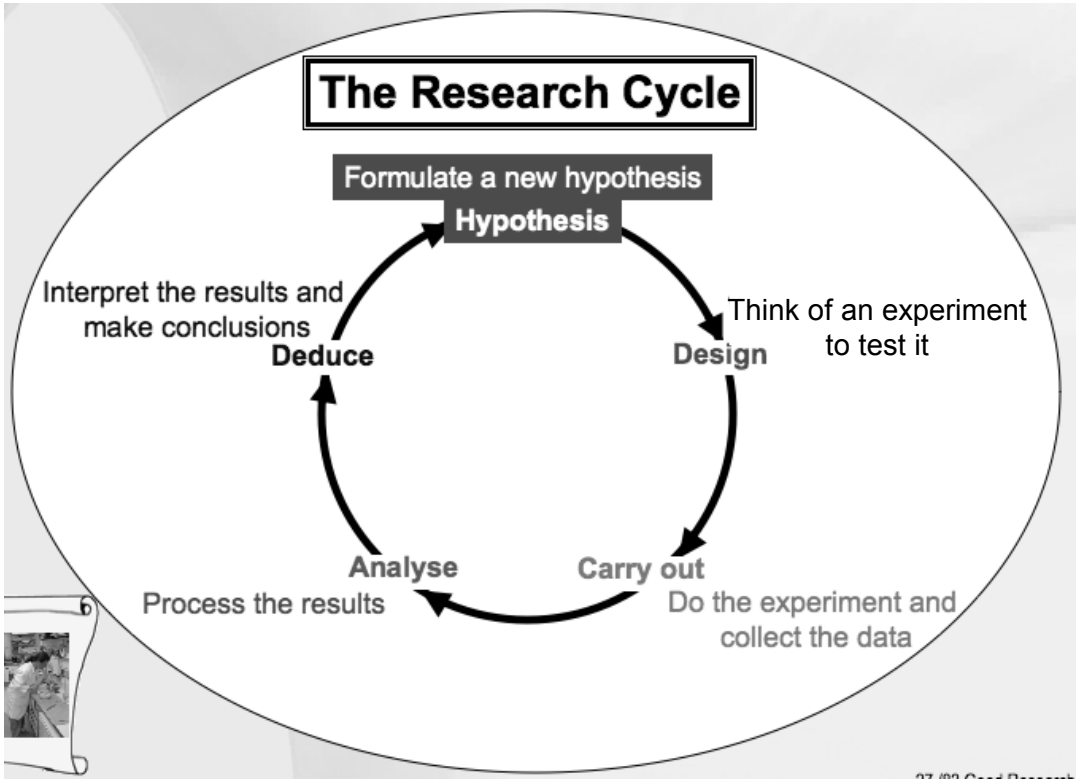
How will you show that the gene expression is due to a change in ABA synthesis and not a direct effect of water stress itself?

The key is designing the experiments to ensure that there can be ***no other explanation*** that would invalidate the test of your hypothesis.

i.e. to tell you the ***truth*** of what happened.

- **Only then will you know whether your hypothesis is *right* or *wrong*!**

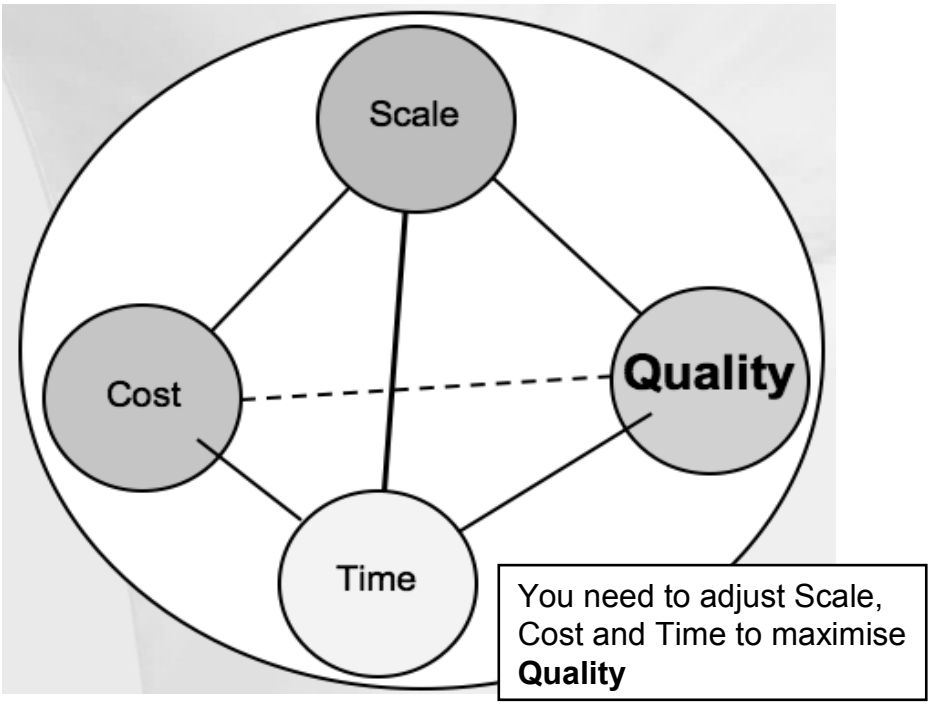
This leads to ...



A good research project also has to take account of four key components:
- the scale of the project
- the cost of the project
- the time available for the project
- the quality of the results.

Each of these factors depends on the others, so they can be considered as a research pyramid …..

The Research Pyramid



Note that the line joining **Quality** to Cost is dashed.
In fact **Quality** rarely depends on Cost!

And the *factor* that influences most the quality of your research is experimental design.

The more time you spend planning your research/experiment, the more likely you are to identify all the factors to control/measure, and therefore the closer you will get to the truth.

Now it's time to get you thinking about identifying factors and setting up hypotheses:

Here's an example that we shall work through together.

On a visit to your local supermarket you see an advertisement for **large decorative sugar crystals [irregular crystals 3-6 mm]** to serve with your filter coffee.

This looks attractive and will impress your dinner guests, so you decide to buy some!
Question: will the large crystals take longer to dissolve than other forms of sugar?
If so, this could be a disadvantage.

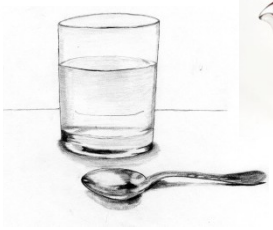
From asking questions you develop hypotheses.

You have some other types of sugar at home, so you decide to design some experiments to test them.

As well as the large decorative sugar crystals [irregular crystals 3-6 mm], in the cupboard you have also found:

- **icing (powdered) sugar**
- **granulated (crystal) sugar**
- **raw cane (brown) sugar**
- **sugar cubes**

In your kitchen, you also have a balance (kitchen scales), a spoon, a glass beaker, a measuring jug, and you have an egg timer and of plenty of course water:


Question: will the large crystals take longer to dissolve than other forms of sugar?
Convert that into a hypothesis …..

What factors are likely to influence the way sugar dissolves in water (given equipment in the last slide)?

Now think up a second hypothesis that could be tested with the five types of sugar.

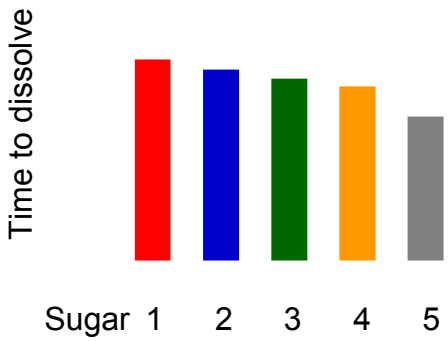
Do you have enough equipment to test your hypotheses?
What other equipment do you need (from your house)?
Do you expect to have any problems that may invalidate the test of the hypotheses?

Discussion of sugar experiments.

So, you have now collected all your data.

Here is an example of how to present the results:

Key: 1 = large crystals, 2 = cubes, 3 = raw sugar, 4 = ordinary, 5 = icing,



- I have assumed that large crystals are slowest to dissolve.

Conclusion from the sugar experiment:

It looks as if the large decorative crystals took longer to dissolve in water than the other forms of sugar.

So, slower to dissolve in your coffee!

Or are they? ... **What is the truth?**

What is missing from the histogram?

Until you know whether differences are ***significant*** or not,
you can't test your hypothesis.

So, how are you going to do this?

Ivo van der Lans will shortly give you advice on this.

Deciding the statistics is where the problems start.

This is why good experimental/research design is vitally important, **as well as good observation skills.**

How do you design your experiments/research to get the best test of your hypotheses?

How are you going to decide on the **factors** to keep constant to ensure the best quality data?

What other **factors** might influence your results that you haven't considered yet?

How many replicate measurements do you need to know whether any differences are significant or not?

Will you use individual samples, pooled samples or can you use paired samples (control - treated)?
What statistical methods do you plan to use to test whether the data sets are different?
Are there alternative statistical methods that would be better **but possible only if you changed the experimental design** (eg paired samples)?

For research which tests the consequences on the subject of study of changing one or more factors, replication is **essential** for good experimental design.

But how many replications are needed for you to find significant differences between two or more datasets?
- this is a question my students have asked me!

[It's a bit like asking the question 'How long is a piece of string?'
Answer - that depends on how long it is!]

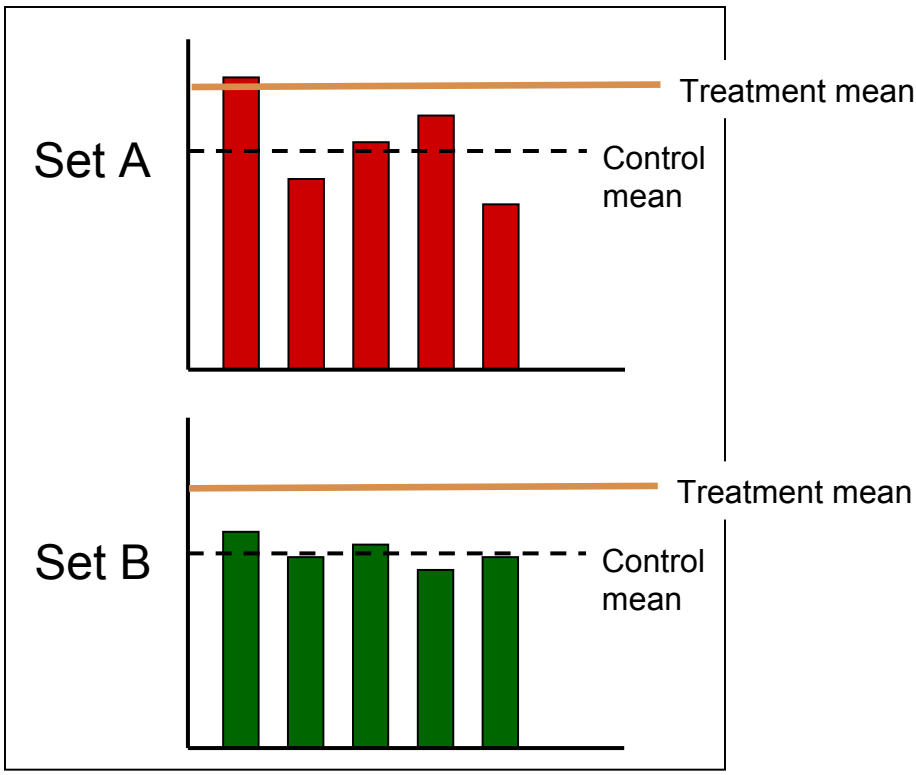
The number of replicates you need will depend on the extent of variation ***within*** a dataset or treatment*.

* A treatment typically refers to one or more ***factors*** which are being modified in the experiment.

The bigger the variation ***within*** a dataset, the more replication you will need in your experiment to find significant differences ***between*** datasets (treatments).

Let's look at an example ….

Take two sets of five replicate values:



It doesn't take a PhD, or even an MSc degree to decide that set A is more variable than set B, so it will need more replicates of set A to find significant differences than with set B.

So, the more success you have in designing the research/experiment to reduce the variation ***within*** a dataset, the easier it will be for you to find significant differences ***between*** datasets.

In the example above, if this was plant height, maybe seed size varies a lot for variety A, but is very uniform in variety B. Therefore, selecting for uniform seeds might help reduce the variation ***within*** the variety.

You need to think about all of this ***before*** you start any experimental/ research work as this will determine your experimental design.

So, the most critical part of the research cycle is experimental (methods) design.
- **get this wrong and you are wasting your time!**

How often have you said to yourselves either during or after a piece of research or experiment:
"I should have thought of that" or **"I forgot to do that"**?

Good research is based on avoiding having to say the comment
"Oh, bother [or suitable local equivalent]! I didn't think of that"
or your supervisor, or worse still, a reviewer of your manuscript saying
"Why didn't you think of that?".

Good experimental design and careful planning **before** starting your research/an experiment will **avoid a lot of pain afterwards**!

Even now, when writing up experiments for publication I sometimes find myself saying "It's a pity I didn't measure that!"

Dragan!

Three pages of extra slides for biologists:
For the biological scientists: Here is a riddle [brain teaser] for you:

When is a significant difference not a significant difference?
Answer - when the variation measured is not due to the test organism/substance.
For example, you want to test whether a particular treatment will affect concentrations of the hormone abscisic acid (ABA).

Now your experiments will have different sources of variation due to:
- effect of a treatment on ABA **which you want to test**, *but also*
- the efficiency of the extraction process for the hormone, *and*
- the efficiency of the method to purify the hormone, *and*
- the reproducibility of the assay to measure ABA concentrations.

Therefore, you need to know how variable are
- the extraction process
- the purification protocol and
- the quantification method

before you can test effects of the treatment on tissue ABA concentrations.

Here's an example of an assay for the plant hormone ABA which was carried out with duplicate analyses of samples extracted from the plant.
If the duplicates differed by more than 10% from the mean, the extract was re-assayed (as in sample 2 - red text).

4 Feb 1999
Assay label: SP5

No.	dpm1	dpm2	Mn dpm	pg/tube	Wt.extr	Ratio	ABA content ng/gDW±se	Sample
1	2512	2514	2513.0	1053.3	38.6	25.9:1	491.1±0.3	169S QTL
2	**2889**	**2412**	**2650.5**	**988.3**	**46.2**	**21.6:1**	**385.0±48.5**	**40N QTL**
3	2515	2575	2545.0	1035.5	31.1	32.2:1	599.3±9.6	57S QTL
4	1750	1740	1745.0	1660.6	37.5	26.7:1	797.0±2.7	148N QTL
5	2876	3141	3008.5	813.8	39.4	25.4:1	371.7±25.5	200N QTL
6	2486	2687	2586.5	1014.4	40.3	24.8:1	453.0±24.3	187N QTL
7	2452	2318	2385.0	1130.5	44.2	22.6:1	460.3±17.0	150N QTL
8	3410	3326	3368.0	673.4	34.3	29.2:1	353.3±7.9	181S QTL
9	2521	2382	2451.5	1090.0	37.5	26.7:1	523.2±19.8	9PN QTL

Look at the consequences for your results of variation by no more than 10% at different stages in the analysis process.

Source of variation	Variation	Type of variation
Experiment to experiment	105-95%	biological
Plant to plant within experiment	105-95%	biological
Extraction efficiency	95-85%	analytical
Purification efficiency	95-85%	analytical
Assay efficiency	105-95%	analytical

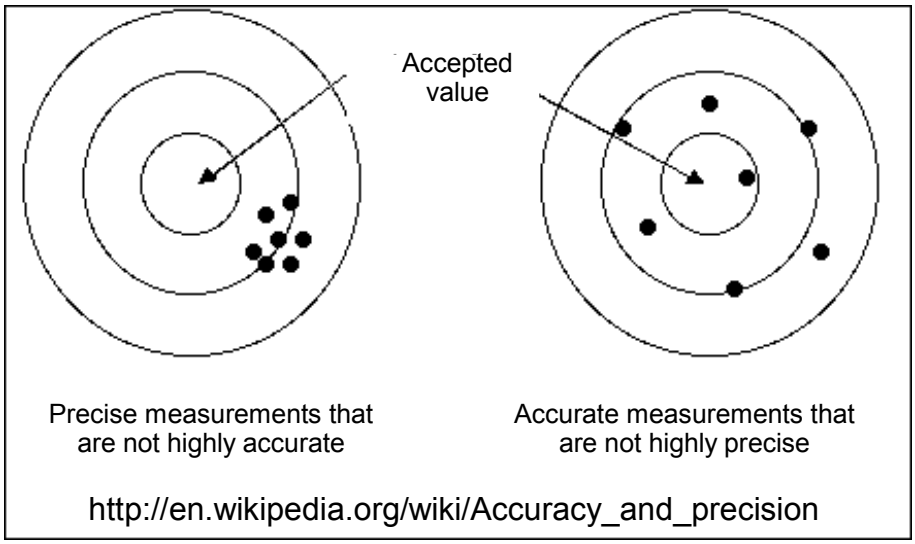
Guess how big you think the range might be overall, if the true value is 100 at 100% analytical efficiency.

A single analysis result could give a value ranging from
100 x 1.05 x 1.05 x 0.95 x 0.95 x 1.05 = 104.5, to
100 x 0.95 x 0.95 x 0.85 x 0.85 x 0.95 = 61.9 !
Biological range = 90.3 - 110.3% of true value.
Analytical range = 68.6% - 94.8% of true value.

Precision and Accuracy

Precision is a measure of how closely the analytical results can be duplicated.
Replicate samples (prepared identically from the same sample) are analyzed to establish the precision of a measurement (of enzyme activity, for example).

Accuracy measures how close to a true or accepted value a measurement lies.
The difference between accuracy and precision is illustrated here:



http://en.wikipedia.org/wiki/Accuracy_and_precision

How to minimise the errors - 1

Good experimental design:

Once you have decided the hypothesis to test, start to design the experiment by trying to identify all the factors that might influence the results (influencing your access to the truth).

Ensure that you have your control and treated samples under identical conditions as far as possible, except for the treatment itself: ensure all factors are equal except for your treatments.

How to minimise the errors - 2

How reliable is the equipment you use to measure things (eg pipettors, pH meters)?

How often is your measuring equipment calibrated for precision and accuracy?
Calibrate your equipment/apparatus:
Make sure it is both precise and accurate …..

How to minimise the errors - 3

With a long-term experiment keep a detailed research diary of ***future*** work to be done to remind you in advance what to do each day

Write up your notes of what you did each day, and **be observant!** Some institutes insist on good notes, independently checked. Make a note of everything (planned or unplanned) that might influence the results of your experiment.

How to minimise the errors - 4

- it is extremely important to think beforehand about the sorts of measurements that ***might*** be useful to take in addition to the planned measurements, and record as many as possible.

- if you make a mistake when writing down a measurement, make sure you correct it in a way that is legible to you and to others

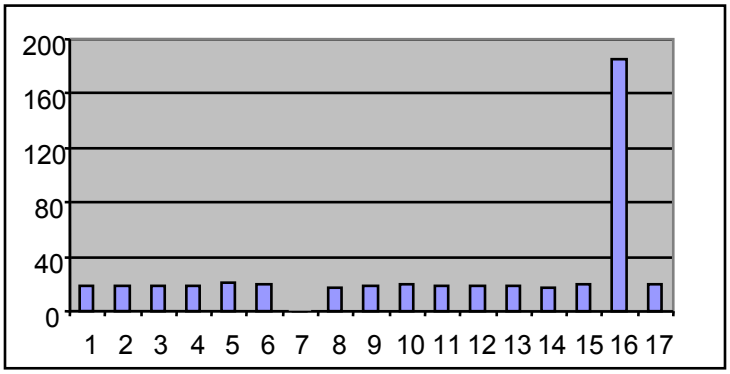
Quality control of your data - 1

- form a judgement beforehand on the sort of mean values that would be realistic for the data set.

- in this way, you are more likely to recognise a value/mean that is a mistake (a number misread when entering into the computer, or a decimal point missed out).

Here's a histogram of a dataset entered into Excel™:

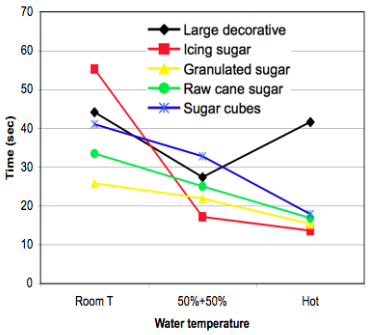
This example illustrates errors with the decimal point: "," instead of "." for no. 7 and "." missed out from no. 16.



Note that errors can very easily occur with Excel™ if you copy and paste or move columns of data that contain a formula and not values.

Here's an example of the effect of poor quality control of data:
- a simulated dataset for the sugar experiment.

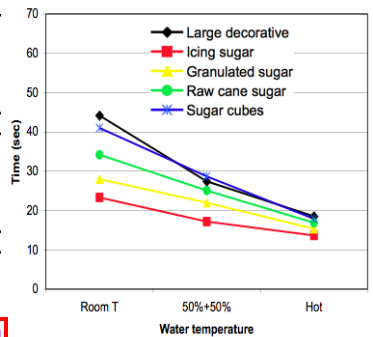
Room temperature	Rep 1	Rep 2	Rep 3	Rep 4	Rep 5
Large decorative crystals	43.2	45.3	44.3	41.2	46.6
Icing (powdered) sugar	23.3	25.4	213	20.1	26.8
Granulated (crystal) sugar	29.4	26.5	27.3	15.2	29.8
Raw cane (brown) sugar	31.2	37.3	30.1	37,5	35.9
Sugar cubes	41.5	43.2	40.6	38	40.4
Hot tap water	Rep 1	Rep 2	Rep 3	Rep 4	Rep 5
Large decorative crystals	19.4	24.1	16.6	154	19.9
Icing (powdered) sugar	12.1	15.9	11.6	15.4	13.3
Granulated (crystal) sugar	16.2	14.4	13.5	16.7	16.1
Raw cane (brown) sugar	17.3	16.1	19.7	14.8	16.4
Sugar cubes	16.8	18.9	19.3	17.2	17.2
50% hot+50% room T	Rep 1	Rep 2	Rep 3	Rep 4	Rep 5
Large decorative crystals	28.3	26.9	29.4	25.1	24.9
Icing (powdered) sugar	18.8	16.9	14.5	19.9	15.8
Granulated (crystal) sugar	22.3	24.3	21.8	20.6	21
Raw cane (brown) sugar	25.3	22.9	26.7	26.2	24.3
Sugar cubes	28.7	29.8	33.1	32.9	35.6



Is this the truth? Have all factors been considered?

Do the results represent only experimental error, or are they experimental + human error?!

Room temperature	Rep 1	Rep 2	Rep 3	Rep 4	Rep 5
Large decorative crystals	43.2	45.3	44.3	41.2	46.6
Icing (powdered) sugar	23.3	25.4	213	20.1	26.8
Granulated (crystal) sugar	29.4	26.5	27.3	15.2	29.8
Raw cane (brown) sugar	31.2	37.3	30.1	37,5	35.9
Sugar cubes	41.5	43.2	40.6	38	40.4
Hot tap water	Rep 1	Rep 2	Rep 3	Rep 4	Rep 5
Large decorative crystals	19.4	24.1	16.6	154	19.9
Icing (powdered) sugar	12.1	15.9	11.6	15.4	13.3
Granulated (crystal) sugar	16.2	14.4	13.5	16.7	16.1
Raw cane (brown) sugar	17.3	16.1	19.7	14.8	16.4
Sugar cubes	16.8	18.9	19.3	17.2	17.2
50% hot+50% room T	Rep 1	Rep 2	Rep 3	Rep 4	Rep 5
Large decorative crystals	28.3	26.9	29.4	25.1	24.9
Icing (powdered) sugar	18.8	16.9	14.5	19.9	15.8
Granulated (crystal) sugar	22.3	24.3	21.8	20.6	21
Raw cane (brown) sugar	25.3	22.9	26.7	26.2	24.3
Sugar cubes	28.7	29.8	33.1	32.9	35.6



The boiler running out of hot water was evidently a **factor**!

What do you do if you don't like the data?

If your experiments do not give you the answers you expect, do not reject the results unless you can prove they are not valid.

Never reject data without a good, a **very** good reason!

Never, **never** invent data, however likely they may be!

Always think about asking yourself the question:

'If I was to stand up in the witness box of a court of law and had to swear on oath to tell the truth, the whole truth and nothing but the truth, would I believe that my conclusions represented the truth of what happened?'

Have I proved my case: guilty or not guilty?

Conclusions:

So, if you think carefully about all the points I've described and spend enough time planning to identify:

- all the **factors** to account for/control in your research
- all the other **factors** that might invalidate your results
- all the comparisons you need to make
- all the degrees of replication that will be needed to test a particular level of significance
- all the quality control necessary of your datasets
- all the sources of error, and how to reduce them
- all the problems and things that *could* go wrong (!),

then ...

you are likely to have a **well-designed experiment** that will give **good quality results**, which will give you

- a valid test of your **hypothesis**
- something worth writing up for one or more good quality **publications**
- a sound basis upon which to develop ideas on what to do next: forming your next **hypothesis**
- but crucially: access to the **TRUTH**
- and that will give you ...
- **satisfaction in a job well done!**