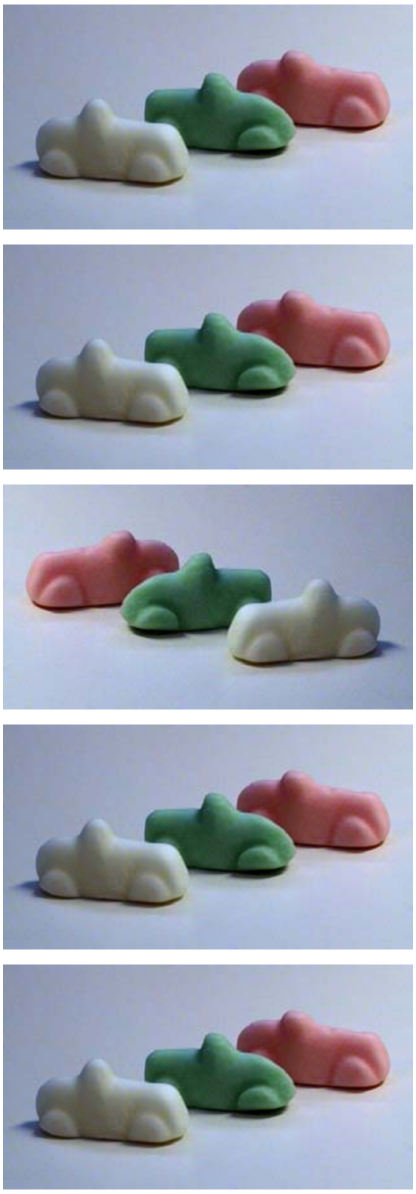
Margareta Wallin, Torgny Bohlin och
Lars Johan Erkell

Department of Zoology
Göteborg University

Educational sweets

How to use sweets to illustrate scientific methodology, statistical methods and scientific misconduct



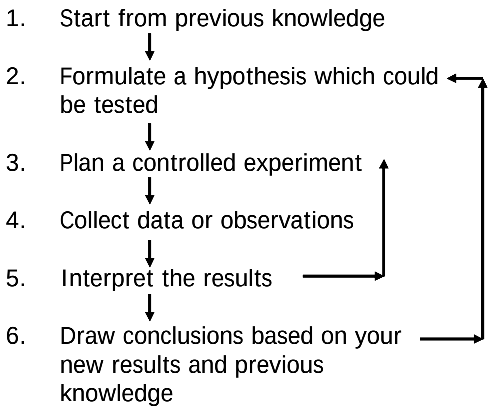
Background to the scientific method

For most people the scientific knowledge we have today is considered as the truth. For a scientific researcher the knowledge of today is based on observations and experiments, but it is an ongoing accumulation of knowledge meaning that it can change.

Scientific knowledge is therefore only as true as the observations or experiments it is based on. It can be modified or changed anytime with results from further research. New experiments can even completely change our view of the world. One such example is that sun energy previously was believed to be the source for all biological life on the earth. This view was abandoned recently when bacteria were found in the deep sea using hydrogen sulphide (H_2S) instead of solar energy. A long time ago we believed that the earth was the centre of Universe. Few people were aware of that proteins could be the cause of illness before the discovery of the Mad Cow Disease. The knowledge could therefore change as soon as we get more information.

The scientific method is based on the generation of a hypothesis taking advantage of previous knowledge. The hypothesis is thereafter tested in many different ways (Figure 1). We can either approve or reject the hypothesis based on our interpretation of the results from well controlled experiments or systematic observations. If our studies are successful (which they not always are), our results will improve current knowledge and perhaps change our view of the world. Previous knowledge is always used both when we formulate our hypothesis and when we present our conclusion. Our results will thereby be used in a larger context.

The scientific method



1. Start from previous knowledge
2. Formulate a hypothesis which could be tested
3. Plan a controlled experiment
4. Collect data or observations
5. Interpret the results
6. Draw conclusions based on your new results and previous knowledge

Facts – the scientific method

This method has been the driving force for the scientific revolution with its start in the fourteenth century. It has completely changed the western society. IT has one important assumption which can be formulated as follows; the scientific method is based on logical conclusions from experiments and observations of the nature and nothing else. This means that conclusions made by scientists should not be influenced by e.g. religion or tradition, but as objective as possible.

The researcher is of course only a human being as others, with traditions, ethics and religious apprehension. The researcher has to deal with ethical questions similarily as other people; is she/he going to take on a well-paid job where new weapons will be developed? Or should she/he take a more uncertain and less well paid job which could be of great benefit for people in the third world; e.g the development of the “Golden rice”. To be able to make such decisions, the researcher can not be objective, but as any human being judge on the basis from hers/his ethics.

The demand for objectivity is a keystone in the methodology of the scientific work. A well known example is the trial against Galileo Galilei (1615). The court of inquisition found him guilty, since “the proposition that the sun is the center of the solar system and does not revolve about the earth is foolish, absurd, false in theology, and heretical, because expressly contrary to Holy Scripture”. The church held the medieval position that the truth, as revealed by the Holy Scripture, had a significance exceeding that all possible experiments or observations that could be made in nature. Galileo held the opposite view. To him, nothing could be closer to truth than reality, and he was not willing (but eventually had to) accept that results from well performed observations and experiments were dismissed with purely theological arguments.

All scientific discoveries are not so well planned as one would believe, a great deal of knowledge is generated by a lucky chance. The discovery of penicillin is a famous example, where the physician Fleming neglected cleaning of his laboratory. This gave rise to the possibility to cure bacterial infections. There are many more such examples, which are not only based on a lucky chance, but on trained researchers able to draw conclusions from their observations putting them into context. Further research in such areas is then performed with usual scientific methods.

Facts – penicillin

Alexander Fleming was a microbiologist at Londons St Mary's Hospital and studied how bacteria grow in petridishes and what factors that could affect the growth. He had found that an enzyme, lysozyme, that is present in tears, inhibits bacterial growth. When he came back from vacation 1928, he started to clean neglected petri ishes that he had left on the bench before his summer vacation. The dishes had got mouldy and most people would have thrown them away. Fleming noted although that no bacteria had grown in the presence of the mould. He understood that the mould had excreted a substance that was inhibitory to bacterial growth. He extracted the substance and it was given the name penicillin. He worked thereafter with scientific methodology; he tested whether penicillin could inhibit growth of other micro-organisms. The development of penicillin for medical purpose could, however, have stopped here, because other researchers were unable to repeat his results when they added the mould *Penicillium notarum* to cultures of staphylococcus bacteria. Many years later it became obvious how the lucky chance had affected Fleming's results, not only by his negligence of cleaning. Ronald Hare, Fleming's assistant, showed in 1964 how Fleming was able to get his results. He showed that optimal growth of bacteria and mould is different at different temperatures. The mould grows best at 20°C while bacterial growth is optimal at 35°C. The colleagues that tried to repeat Fleming's results failed because they cultured the petridishes at 35°C. But how did it happen in Fleming's dishes? Well, he had forgotten the petridishes at the bench and this specific summer when he was on vacation there was nine days with exceptionally cold weather followed by hot days. The mould grew well during the cold days and produced penicillin and when the temperature increased the growth of bacteria increased except where the concentration of penicillin was high. The mould was also of a very special kind that produced very high concentrations of penicillin. It turned out that they came via air from the laboratory at the floor below where a collegue to Fleming worked with them.

Practical work for the classroom

How can you in a simple and fun way get your pupils to understand the research method, why and how statistical methods are used and what is meant with research ethics? We present here one such example fulfilling the demands above in a tasty way. It can be used with a science lecture with or without collaboration with mathematics. It can be used for several different ages depending on how advanced the statistical part is made. In this example we have used satchels with Ahlgren's car but many other kinds of sweet can be used as well. In Great Britain we have used satchels with M&M.

Observation: We have bought a satchel of Ahlgren´s cars with green, pink and white cars. The number of green cars dominates in the satchel over white or pink cars. Since we prefer green cars we by another satchel, and green cars dominates here as well. We suspect that green cars always dominate after a quick look on the shelf with Ahlgren's cars in the supermarket. Our hypothesis is therefore that the green colour is dominating in the satchels. With other words, this is our view: Satchels where green cars are dominating are more common than satchels with pink or white cars.

How do we test our hypothesis with scientific methodology? Well, it is not enough that we determine the numbers of green, white and pink cars in one or two satchels to be certain that this is general for all satchels. It would be like interviewing one or two electors to be able to make a statement about the political sympathies of a country. We have therefore to buy a number (N) of satchels, and to count the number of green, pink and white cars in every satchel (in an electorinvestigation N is usually one or two thousands of people). The number of satchels we use is called a sample test. Make a table for the first satchel. The result could be the following:

Number of **green** cars	Number of **pink** cars	Number of **white** cars	More **green** than **pink** or **white** cars YES/NO
38	27	34	YES

Do the same with all satchels. Let's say that we will investigate a sample test of N = 10 satchels and we get the result of 7 YES and 3 NO; *e.g.*, 70% against 30%. Is it possible to draw the conclusion that green cars are dominating in the satchels?

No, you could never be completely sure unless you count every satchel that is produced by the company. This is of course impossible. But if we had got the result 10 YES and 0 NO, we would most probably had stronger reasons to believe that our hypothesis was right that if we had got 5 YES and 5 NO (or if we got fewer YES than NO).

This discussion illustrates the basic principle for a statistical test. We determine in some how probable an observation is (in our example 7 YES and 3 NO) if it in reality is no difference. If green cars are not dominating, how probable is it that we get 7 YES and 3 NO? This probability is usually called p. It is improbable that green cars is not dominating if p is low - which means that it is probable that green cars are dominating (those who have some mathematical knowledge might understand that this test uses binomial distribution, but this is not anything the rest of us needs to care about). We can get p-values directly from the table below. N = the size of the sample test (in our example 10). The value k in the table is the lower of the number of YES and NO, in our case k = 3. We can read p = 0.172 in the table where N = 10 and k = 3.

What does this mean? It does mean that it is not unusual to get the occurrence 7 against 3, even if green cars do not dominate. It happens with the probability of p = 0.172, meaning that it occurs 17 times out of 100 times. If we get 7 YES and 3 NO we cannot draw the conclusion that green cars dominate with reasonable certainty. What conclusion can we draw if we had got 9 YES and 1 NO (N = 10, k = 1)? Look in the table and you will find that p = 0.011. The probability that green cars do not dominate is only about 1%. With 9 YES and

1 NO there is a strong certainty for our hypothesis that green cars dominate.

What is then “a strong certainty”? It has become a convention in many research areas to set a limit at p = 0.05. If one gets a value on p which is 0.05 or less, the hypothesis that there is no difference can be rejected and the hypothesis of that there is a difference can be accepted. If p is larger that 0.05 the support for the hypothesis that there is no difference is so big that one cannot reject it.

In our example above with 7 YES and 3 NO we cannot accept the hypothesis that green cars dominate. We could however accept it if we got 9 YES and 1 NO. The smaller the p is, the more likely it is that we draw the right conclusions.

One could also start by using the table. Let us assume that we can afford to by 15 satchels with cars. Suppose that we use p = 0.05 as a limit for decision (the significance level). How many NO of these 15 satchels can we get at maximum and still be able to draw the conclusion that green cars dominate? Look at the significance figures in the table with N = 15 and read k in the column where p is 0.05 or less. It turns out that k = 3 (p = 0.018). (If you read the value for p at k = 4 you will find that it is 0.059 which is bigger than the decision limit we have chosen). The occurrence of 12 YES and 3 NO would support our hypothesis that green cars dominate, but 11 YES and 4 NO would not do that. (If you get the occurrence 11 YES and 4 NO and you really want to solve the question, you have to try to be able to buy more satchels).

Discuss in the classroom why you have to use statistical methods in science and medical studies, and in which cases you do not need to do it or are unable to do it.

Table of probability

N	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
4	0.062	0.312	0.688	0.938	1.0											
5	0.031	0.188	0.500	0.812	0.969	1.0										
6	0.016	0.109	0.344	0.656	0.891	0.984	1.0									
7	0.008	0.062	0.227	0.500	0.773	0.938	0.992	1.0								
8	0.004	0.035	0.145	0.363	0.637	0.855	0.965	0.996	1.0							
9	0.002	0.020	0.090	0.254	0.500	0.746	0.910	0.980	0.998	1.0						
10	0.001	0.011	0.055	0.172	0.377	0.623	0.828	0.945	0.989	0.999	1.0					
11		0.006	0.033	0.113	0.274	0.500	0.726	0.887	0.967	0.994	0.999+	1.0				
12		0.003	0.019	0.073	0.194	0.387	0.613	0.806	0.927	0.981	0.997	0.999+	1.0			
13		0.002	0.011	0.046	0.133	0.291	0.500	0.709	0.867	0.954	0.989	0.998	0.999+	1.0		
14		0.001	0.006	0.029	0.090	0.212	0.395	0.605	0.788	0.910	0.971	0.994	0.999	0.999+	1.0	
15			0.004	0.018	0.059	0.151	0.304	0.500	0.696	0.849	0.941	0.982	0.996	0.999+	0.999+	1.0
16			0.002	0.011	0.038	0.105	0.227	0.402	0.598	0.773	0.895	0.962	0.989	0.989	0.999+	0.999+
17			0.001	0.006	0.025	0.072	0.166	0.315	0.500	0.685	0.834	0.928	0.975	0.994	0.999	0.999+
18			0.001	0.004	0.015	0.048	0.119	0.240	0.407	0.593	0.760	0.881	0.952	0.985	0.996	0.999
19				0.002	0.010	0.032	0.084	0.180	0.324	0.500	0.676	0.820	0.916	0.968	0.990	0.998
20				0.001	0.006	0.021	0.058	0.132	0.252	0.412	0.588	0.748	0.868	0.942	0.979	0.994
21				0.001	0.004	0.013	0.039	0.095	0.192	0.332	0.500	0.668	0.808	0.905	0.961	0.987
22					0.002	0.008	0.026	0.067	0.143	0.262	0.416	0.584	0.738	0.857	0.933	0.974
23					0.001	0.005	0.017	0.047	0.105	0.202	0.339	0.500	0.661	0.798	0.895	0.953
24					0.001	0.003	0.011	0.032	0.076	0.154	0.271	0.419	0.581	0.729	0.846	0.924
25						0.002	0.007	0.022	0.054	0.115	0.212	0.345	0.500	0.655	0.788	0.885
26						0.001	0.005	0.014	0.038	0.084	0.163	0.279	0.423	0.577	0.721	0.837
27						0.001	0.003	0.010	0.026	0.061	0.124	0.221	0.351	0.500	0.649	0.779
28							0.002	0.006	0.018	0.044	0.092	0.172	0.286	0.425	0.575	0.714
29							0.001	0.004	0.012	0.031	0.068	0.132	0.229	0.356	0.500	0.644
30							0.001	0.003	0.008	0.021	0.049	0.100	0.181	0.292	0.428	0.572
31								0.002	0.005	0.015	0.035	0.075	0.141	0.237	0.360	0.500
32								0.001	0.004	0.010	0.025	0.055	0.108	0.189	0.298	0.430

k

Values less than 0.0005 are not included

How you discuss and illustrate fraud in science in the classroom

When you have used the sweets in the classroom to illustrate the scientific method and statistical methods, you can use the same material to demonstrate what is meant by fraud in science (or scientific misconduct). You could present how many green, pink and white cars you found in a satchel where you have eaten several pink or white cars. The result from this satchel will thereby look different. How big is the probability that anyone in your class would suspect that you have changed the results by eating some of the cars? You could discuss with your class the unwritten rules that exist in science that every researcher should clearly state why the study was done, how it was done, which results that came out of the study and the meaning of the results. The scientific work has to be done carefully in such a way that it can be repeated by other scientists (*e.g.*, Fleming's case where it took a long time before the results could be repeated and accepted by the scientific society).

A researcher could of course get results that support the hypothesis that green cars dominate, by eating several pink or white cars if the observations seem to reject the hypothesis. Another possibility for scientific misconduct is that the researcher decides to eliminate a satchel filled with mainly pink cars with the motivation that it is not representative for satchels in general since it is so different from the others.

Are there really researchers that perform scientific misconduct or perform their work in a careless manner? If so, can we trust scientific data at all? Well, a minority of researchers fall into this category. The reasons for the misbehaviour could be several; it could be personal characteristics, but also caused by the wish to publish much research results in order to get top positions and more research funding, or even to keep the position and research funds. A researcher could also have an interest to manipulate the studies for political and commercial reasons, which could be beneficial for the researcher.

The scientific community has several mechanisms for self-correction. Before and after publication, all work is scrutinized by colleagues within the same field of research. Also, the subject science ethics is growing and many universities have today ethical advisory boards and education of staff and students to be able to prevent scientific misconduct. No doubt, scientific misconduct and careless research generate some friction in the scientific machinery, but the machinery still works.

Referenses:

Ennos, R. (2000) *Statistical and Data Handling Skills in Biology.* Prentice Hall. ISBN: 0 582 31278 7.

Forsman, B. (1997) *Forskningsetik. En introduktion.* Studentlitteratur. ISBN 91 44 00484 2.

Le Fanu, J. (2000) *The rise and fall of modern medicine.* Carroll and Graf Publishers: Inc. New York. ISBN 0 7867 0732 1

Resnik, D.B. (1998) *The Ethics of Science: An Introduction.* Routledge: London/New York. ISBN 0-415-16698-5.

Thompson, J., Baird, P. & Downie, J. (2001) *The Olivieri Report.* Lorimer: Toronto. ISBN 1 55028 739 7.
(This book is about the Canadian physician who fell into disgrace when she wanted to publish that the drugs she was testing for a pharmaceutical company had serious drawbacks. Her case has inspired John Le Carré to write the thrilling book “The Constant Gardener”).

www.ahlgrensbilar.com

Photo:

Dean Madden, National Centre for Biotechnology Education,