A proposal for testing soft versus strong poison.

Reinko Venema.

Introduction. Recently in China thousands and thousands of babies got very sick because of illegal additives in the milk. The scale and size of this corruption are unprecedented as far as I know, rather likely it is done by people who could not oversee the damage of their short sighted actions.
Since it is in the news I decided to write down the statistical lessons that I gave to students in the city of Utrecht in the years 2000 and 2001.
My thanks go to professor Jaap Stam, from him I learned how to handle things like testing on poison. From him I learned what to put in your zero hypothesis and what not.
I will try to make this as non technical as possible, so I skip all that t-test stuff.

How statistical testing works. A very common way is writing down a so called zero hypothesis against a denial of that zero hypothesis, it is important to understand we do not know what the real state of the world is and we have only some statistical sample to try to guess where we are.
Often it looks a bit like this:
H_0 In theory, this is reality.
H_1 This is the denial of reality.

After having crafted the zero hypothesis and it's denial we always calculate from the zero hypothesis, that is we accept that one for being true and we look at our statistical sample results and we calculate the likelihood of our statistical sample or worse.
We statistical folks create a so called statistic and we calculate the tail possibility of our statistical result, if the result is too small we reject the zero hypothesis.
Simple example: We throw a coin one hundred times to see if the coin is unbiased, that is our sample.
When the results are 70 times head and 30 times cross we calculate P(cross is 30 or lower) or P(heads is 70 or more). When calculating this we use as zero hypothesis:

$H_0 : \pi = 0.5$ versus
$H_1 : \pi \neq 0.5$

In fact the chances of a tail event like 70 heads or more is so small that we reject the zero hypothesis (the unbiased coin) and we accept the alternative that the coin is biased.

Now we fly the wisdom of professor Stam into the building: But in practice when you craft your hypothesizes, there are all kinds of costs when you make the wrong decision. For example the statistical department says 'The food is not ok' in that case the production is stopped and there are all kinds of costs to that. When the statistical department says 'The food is ok' and later it turns out it is not, there are also huge costs involved.

How do you walk the fine line on that?

Two theoretical examples on walking fine lines. Example one:
There is a large industrial laundry that takes in the dirty stuff from a few hospitals in the neighborhood. The laundry has a lot of phosphorous related products in her waste water that is delivered to the lakes around. In the past the phosphor related stuff was so bad that most of the fish and pants died in those lakes. Therefore after waiting many years the local government decided to put some threshold values on the poison the laundry could pump away.

Example two:
In a densely population as we have here in Holland, you cannot bury the waste that the consumers produce. You have to destroy it. One of the ways of destroying the waste is by burning it in large ovens. The large ovens expose their internal wisdom to the skies and you are allowed to breathe that wisdom in.
The problem is: When garbage burning factories do not burn at a high enough temperature there will be lots of doixines produced. This poison will be spread around and the local cows will eat grass spread with this dioxin tainted smoke.

How do you walk the fine line on this?

What are the costs in both examples? In example one when the laundry is allowed to produce while the phosphor levels are too high, the costs are worse water quality in the lakes. When wrongfully the laundry cannot work any longer because the statistical test is 'too sharp' the hospitals cannot work properly, may be operations are delayed and stuff like that.

In example two when batches of milk are allowed that contain too much dioxin, this is bad for the health of the general population because dioxin is a very toxic thing. When wrongfully the batch of milk is rejected the farmer does not make any money and may be the farmer goes to the courthouses in order to place the damage on the waste burning companies.

In example one we give the laundry the benefit of the doubt, we work from the assumption that they are responsible people. That means that only when phosphor levels are far to high, the production will be stopped.
In example two we do not take any risk and we work from the assumption that the dioxin levels are too high anyway and only when the samples say it is 'low enough' the milk is allowed to be processed.

How to formulate your zero hypothesis? For the laundry it looks a bit like this:
H_0 : The phosphor levels are likely ok,
H_1 : The phosphor levels are too high.

For the milk factory it will be as next:
H_0 : The dioxin levels are too high,
H_1 : The milk can be accepted because it has been proven it is low enough to be accepted.

If μ represents the theoretical average level of the poison involved and μ_c denotes the 'critical threshold' this looks like the next for example one:
$H_0 : \mu \leq \mu_c$
$H_1 : \mu > \mu_c$

And for example two where we really don't want to take any kind of risk with accidentally allowing to high levels of dioxin this is:
$H_0 : \mu \geq \mu_c$
$H_1 : \mu < \mu_c$

The difference is very easy to explain:
In example one you put the laundry out of service when phosphor levels are too high.
In example two the milk factory can only go to work if dioxin levels are low enough.

So far the theory; in theory life is simple, in practice it is complicated because you rely on lab results. If they make mistakes in the lab your model might be perfect but you will make faulty decisions anyway. And as always you have to guard against corruption: when batches of milk are changed before they go to a proper working lab again you will make the wrong decision of allowing milk that should have been denied to enter the food chain on the consumer level.

As a wise man once stated: You can use a model that points you towards some risks, but that will not prevent you from the risks of using that model...

Epilogue. The above model was used in the exams of a lot of students in the year 2000 or 2001 in some business school in Utrecht, the Netherlands. A few months later I was sitting in the train to go to work in Utrecht and I was reading a newspaper. To my amazement there was a news article that stated that working standards for measuring dioxin in milk was changed a little bit, according to an 'expert' when using this new standard a few batches of milk that were allowed in the past would have been rejected in the future. Of course there was 'no harm' for the population in the past but now the standard was the way it was supposed to be anyway...

With a little smile on my face I poured in another cup of coffee from my thermos bottle and added some milk...