Lies, damned lies, and legal truths

Richard Gill
Mathematical Institute
Leiden University

http://www.math.leidenuniv.nl/~gill

My acquaintance with Aernout Schmidt began when we were asked to be one another's opponent in a Leiden science-café debate on the celebrated case of the Dutch nurse Lucia de Berk, who at the time was serving a life sentence for seven murders and three attempted murders of her patients: children at a special children's hospital, and terminally ill old people in an ordinary hospital ward where she had earlier worked. The case was sparked when, on the early hours of 4 September 2001, for the so-many'th time (as it appeared) a young child died during one of her shifts. The statistical question of whether Lucia's repeated presence at a series of deaths and near-deaths could merely have been a coincidence was answered first for hospital authorities, then for police investigators, and finally in court (in 2003; answer: no, it could not have been chance).

Aernout is a specialist on law and information technology, and in this capacity he has even taught statistics to lawyers. I am a statistician, but infamous in some Dutch legal circles for my part in inciting some kind of mass movement to get Lucia a re-trial, even though the verdict had been confirmed first on appeal (2004) and finally on "cassation" at the Supreme Court (2006). For instance, G. van Manen, and later P. J. van Koppen (the latter a law-psychologist and eloquent criticaster of our judicial system), accused me in Nederlandse Juristenblad, The Journal of Dutch Lawyers, 2008, of deliberately playing to the gut-feelings of an international scientific rabble by feeding them with misinformation and lies (shades of the lynching of the brothers de Witt).

A retrial is presently (2009) underway following identification of new facts in a medical investigation commissioned by the Procurator-General to the Supreme Court (2008).

Unfortunately for the heat of the science-café debate, it turned out that Aernout and I got on very well together, and in particular tended to agree on almost everything. Still it was a lively and exciting evening and promised well for future collaboration. We quickly converted our debate into a short paper in the new Dutch journal "Expertise and Law" (2008) and I was hoping for more collaboration in the future, only to be shocked by the information that Aernout was so much older than he seems that his friends were writing him a Liber Amicorum on the occasion of his retirement. Being asked to contribute I eagerly sieze the opportunity to settle one niggling difference with Aernout, and to fulfil one wish of the editors of "Expertise and Law", which we did not do in the paper.

This paper will discuss in essay form (more precisely, a Joycean stream of consciousness) my difference with Aernout, and intertwined with that, supply the "missing" passages of our paper. The two topics are indeed in my opinion strongly linked.

Our difference of opinion concerned the question of whether statistics still plays any role in the Lucia case after the 2004 appeal, at the conclusion of which the judges explicitly wrote that their verdict (Lucia is guilty of seven murders and three attempted murders) is reached without any use of statistics.

The (deliberately) missing passages of our paper concerns the question what kind of statistics, if any, ought to be used in criminal court procedings: Classical/Frequentistic; Full Bayesian; or according to the latest fashion: the Likelihood Ratio. We only briefly mentioned the first two statistical flavours, and totally ignored the new *flavour of the month*.

Was Lucia convicted by statistics?

So, is statistics still present and still important in Lucia's case, after the appeal? My opinion is that despite the judges' words (well – that depends on how one reads them), the written argument for their verdict is statistical through and through: however, unfortunately, based on *wrong* statistics, *wrongly* analysed, and *wrongly* interpreted. You may not agree with me, as Aernout also did not agree (though he agreed he should study the verdict first in detail); but it also became clear that he does not understand the word statistics as I do.

I have to admit that my claim also depends on information in the dossier but not in the written verdict. If you wish to check my claims you will be forced to trust the information about it which can be found in Ton Derksen's splendid (2006) book: *Lucia de B: Reconstruction of a Miscarriage of Justice* (unfortunately still not translated in its entirety into English, though some sections and summaries have appeared as separate papers).

The important point I want to make is that there is *no* clear boundary-line between statistics as an advanced and complex science, and common sense interpretation of "statistical data". More subtly, when does the interpretation of statistical data concerning medical events belong to statistics, and when does it belong to medicine? This leads to dangerous situations from the point of view of *determining the true facts of the matter*, which is the task of the judges in Dutch criminal proceedings. Of course, judges often take recourse to the interpretation by experts of "hard facts", when the implications of those hard facts are difficult or impossible for a layperson to see. But what if the court is not aware that certain hard facts do need expert interpretation, because a lay person's interpretation can be easily wrong, though at face value it might seem convincing?

I am not going to propose a solution to this problem, and probably from a legal point of view it does not exist. The law is the law, judges act on it. However if the problem is associated with recurring miscarriages of justice, then there is a long term problem for justice and for society, as well as the damage done in the short term to people's lives.

The close to 100 page verdict after the appeal proceedings (I suppose the longest in Dutch criminal law history) starts in its preamble with the statement that "a statistical probability calculation plays no role at all in our deliberations". Moreover, "every single one of the deaths (and other incidents) has been indisputibly proven by medical-scientific evidence to be unnatural". These are powerful claims and from their prominent position in the preamble to the "arrest" they seem to be aimed at the world at large, in particular at the media and at the scientific community, in which a great debate had raged over the question of whether or not statistics could be used to prove that someone is a murderer; the debate further muddied by strident fundamentalists shouting which kind of statistics should be used (frequentist or Bayesian). From that verdict onwards, whenever a statistician like myself said something in public about the case, the retort was "but statistics no longer plays any role in the case, so you should keep your mouth shut".

My claim is that statistics was almost the only evidence against Lucia. By the device of converting statistical arguments either into common sense arguments of lawyers, or into

medical arguments of medical doctors, the statistics of a coincidence was/were disguised as indisputable medical-scientific facts of unnatural deaths.

Naturally there is a grey area where one person's scientific expertise becomes another person's common sense. This is partly a question of demarcation and ownership of concepts and knowledge. From a scientific point of view, demarcation is merely a matter of convenience and culture; ownership does not exist. For many lawyers however, for whom (to people like me) form often appears to take priority over content, demarcation and ownership are crucial. Obviously, only a statistician is qualified to make statistical probability calculations concerning exactly how unlikely it is that one particular nurse is present at each and every one of the only 9 deaths and reanimations on her ward in a particular year; but a coincidence like this remains (appearing to be) an extraordinary coincidence, and it remains shouting to be given an explanation. When a medical doctor tells you that these events took place on a ward where normally there were no deaths at all (and supports this by showing the numbers of zero deaths in each of the two preceding years), and reminds you that this was a medium-care ward where all the patients were all expected to shortly go back home! – well, who needs a statistician?

Especially, who needs a statistician, when statisticians clearly disagree on what the probability is, because of incomprehensible disagreements concerning the right "model" and the right "paradigma" (these are different issues; the latter refers to the frequentist/ Bayesian debate). Indeed, the defence introduced a probabilist with a philosophical bent, and an expert on artificial intelligence and logic, to argue that were many different probabilities depending on your model, so that the question asked by the court "how likely is it that the coincidence is just due to chance" *could not be answered*. The defence got its way. In its verdict, the court scrupulously avoiding any probability calculations, and even any use of words like statistics.

I will briefly expose and analyse the hidden statistical arguments of the court, and (briefly) contrast them with what I believe are sound statistical arguments. Common sense is not always right in matters of probability and statistics; medical arguments about statistics are very often flawed; and as for legal arguments about statistics ... it is not for nothing that the two most famous misinterpretations of conditional probabilities are called the *prosecutor's fallacy* and the *defence attorney's fallacy* respectively. Please note: I do not want to argue that Lucia was unjustly convicted. The judges were totally convinced of her guilt, and whether or not their written legal argument holds scientific water might be a subsidiary issue for some; especially now that new facts (seem to) have made it redundant. No: the point is that there will be problem in legal proceedings when common sense and scientific expertise have a large overlap; an even greater problem when several expertises overlap with one another and with common sense. In the Lucia case I certainly believe that "law" abused "statistics" in an appalling way. But the blame can be placed just as firmly on the statistical community for being such a shadowy presence in modern Dutch society that the legal community hardly knows that "we" exist, and certainly only has the vaguest idea of what we are about. The incompatibility between the statistical mind and the medical mind is well known to all who work in the arena where those disciplines overlap. Probably the incompatibility between legal and statistical minds is even greater.

I think that the Lucia case has stirred scientific opinion in the Netherlands so strongly, that for some time to come, scientists will be taking more interest in law. This cannot but serve both justice (in its human or social sense) and the legal system. It's an ill wind which blows no-one any good.

R.D. Gill, Lies, damned lies, and legal truths (2010), pp. 39–50 in: L. Mommers, H. Franken, J. van den Herik, F. van der Klaauw and G.J. Zwenne, Het Binnenste Buiten (Liber Amicorum ter Gelegenheid van het Emeritaat van Aernout Schmidt), eLaw@Leiden, Law Faculty, University Leiden.

Statistical evidence

Statistical evidence is not in principle different from any other kind of technical or forensic or scientific evidence. By definition, scientific evidence in the legal setting is evidence which an average person or an average lawyer cannot safely interpret on their own, but for which they have to rely on the expertise of a specialist. Statistical evidence *is* scientific evidence. However, whereas the average lawyer probably realises that she is no expert on ballistics, or accountancy, or toxicology, and will readily acknowledge that there do exist reliable and highly-trained persons who can help them out in these fields, the average lawyer, just like the average human being, is a biological machine whose survival and evolution revolves around recognising and interpreting coincidences (what else is statistics?) and, incidentally, around recognising and interpreting motives and personalities (what else is psychology?). Thus a huge part of the evidence evaluated and sifted by police investigators, lawyers, and judges, is in a very strong sense statistical evidence and/ or psychological evidence. In many situations, one can make do with the statistics and psychology "of the man in the street". A verdict of a criminal court ought to convince any reasonably intelligent and well-thinking person. How to trust statistical conclusions which only a statistician can obtain?

The statistics of the Lucia case are complicated and subtle and still not understood; in fact, largely because of the refusal of the hospital authorities to let any outsider get any look at all at original data. To this day, not a single professional statistician has ever been officially involved in any of the trials and retrials. (If you are familiar with this case, and you are a Dutch lawyer, this statement is meant to tell you that my definition of "statistician" is different from yours). By the way, similarly, to this day, no-one with a broad generalistic medical knowledge has looked at the medical aspects of the whole case in depth, except for the chief paediatrician(or chef de clinique) of the Juliana Kinderziekenhuis, who was a key player in the initial chain of events triggering the police investigation, and her sister-in-law medical doctor Metta de Noo-Derksen (specialist in gerontology and nursing), a key player in the extra-legal investigations which finally led to Ton Derksen's book, and from there by further steps presently irrelevent to my story, to a re-trial.

In my opinion, statistics (and psychology) drove the case, from start to end: Lucia's conviction for serial murder, and even the "the proof" that there were any murders at all – let alone by whom – were almost entirely based on wrong statistical data, wrongly analysed, and wrongly interpreted; by amateurs.

I summarize my "hard evidence" for this at the end of the paper.

Which statistical paradigma?

The second issue discussed in this paper is the sticky question of which statistical paradigma might be appropriate to the Lucia case in particular, and to legal proceedings in general. Aernout and I briefly discussed the opposition between classical frequentist statistics (p-values, hypotheses tests, signficance and power…) and those magically appealing Bayesian statistics (so simple: posterior odds equals prior odds times likelihood ratio). The reviewers of our paper were highly disappointed that we did not pay any attention to a very modern alternative; I think they were referring to the new dogma that all a statistician must do is compute the likelihood ratio. I deliberately did not want to discuss

competing schools of statistics in the context of our debate, nor of our paper, for the following reason: in my opinion, it was totally irrelevant.

It should not be a surprise that there are different ways to formally model the processes of learning from uncertain data, decision making in the face of uncertainty, statistical inference, and so on. There are so many different kinds of ways in which we need to draw conclusions from "statistical data", in the sense of data which has been produced by some random mechanism (I mean, the output of the mechanism is random), and could easily have been different (our task being to draw conclusions about the nature of the mechanism). For instance, when one spins (rather than tosses) a coin, it turns out that coins have a preference to fall on one particular side. This is because coins are struck from a flat sheet of metal, with, say, heads on top, tails below; this procedure makes coins in the shape of a thin slice from of a very long tapering cone. Küchenhoff (2008) reports 501 tails in 800 spins of recent German 2 Euro coins (16 students each spinning a coin 50 times). He noted that there is a significant difference between the students (or their coins, or the tables on which the coins spun, or whatever...).

Is the aim merely to report some kind of fair summary of the data which anyone could use for their own ends? Or should one take account of the use which the consumer is going to make of your statistics? For instance, send someone to jail for counterfeiting coins of the Euro-realm? Should we make use of other information and if so, what and how? Results of spinning 1 Euro coins? Or of Dutch 2 Euro coins? Complex phenomena need to be simplified before they are amenable to mathematical analysis. The different schools of statistical inference correspond to different idealisations; different simplifications. Each one can be extremely appropriate or can be extremely *inappropriate* in any particular situation – it depends on whether one has focussed on essential or on peripheral issues; whether one has neglected essential aspects of the problem en concentrated on peripheral. It depends on how seriously one should take the various kinds of information which are available. And of course, if we are talking about chances, it depends on which chance you are talking about. A chance of 1 in a 100 means 1 in a hundred somethings, and those somethings need to be extremely carefully specified. The same event has any probability between 0% and 100% you like, as one considers it relative to different "classes of hypothetical repetitions".

The debate which raged between Bayesians and frequentists concerning the Lucia case, and which probably put the court off making any serious use of statistics, was largely irrelevant since the crucial thing was not the 1) style of inference, but 2) the model assumptions and 3) the data. Point 2) concerns the question *what are the conceptual repetitions*? What structure can be assumed about those repetitions? This is a very difficult task: we are talking about a conceptual or counter-factual situation in which we rerun a year in the life of a hospital ward again and again, with some things kept the same, other things allowed to vary naturally. Which should be thought to be kept fixed, which are allowed to vary? How will they vary, then? What is *natural variation* in a year in the life of a hospital ward?

What statistical model?

If we want to know if Lucia's presence at so many incidents was due to chance, we need to understand how shifts of nurses, and incidents in the medical care of patients, get fixed in time, in the *natural situation* that no killer is at work; it is just business as usual. The statistician whose advice to the court made a great deal of impact in the original trial

thought of the normal situation as being as follows: first of all, we fix the shifts of all the nurses. Think of a duty-roster being drawn up on the 1st of January, and then being followed scrupulously throughout the year. Three shifts a day, the year round, 365 days; somewhat more than 1000 shifts altogether. During the year patients are admitted to the ward and discharged from the ward. Occasionally there is an *incident* – a life-threatening, and sometimes life-terminating, medical emergency. These incidents were supposed to occur *completely (or uniformly) at random*, in the very strong sense that each and every one of those 1000+ shifts has the *same* chance to have an incident occur in it; and the occurrence or non-occurrence of incidents in any particular set of shifts, has no effect whatever on the chances of incidents in the other shifts.

I think it should surprise no-one that this model turned out to be obviously false. To begin with, the statistics were only being done because a nurse had already noticed that Lucia had been often present at incidents.

What statistical data?

 I will come back to the model later, and say a little more about the data now. An extraordinary fact is that no "statistician" officially involved in the case so far went back to the sources (ward log books, personnel records, patients medical records, and the memory of nurses and doctors) but all assumed that it was valid. I have consulted with statisticians with large experience of similar cases in the UK and in the USA and in Norway. The first thing you do is ask for the formal definition of "incident" and the formal definition of "the shift in which the incident occurred". You check that the criteria can be verified, for every shift, without having to know which nurses were on duty when. Similarly you ask for the formal definition of "who's on a shift" and you check that the criteria can be verified without knowing what happened during the shift, for every shift.

Another fact is that the tabulation of this data was done largely by an internal team of medical personnel at the hospital where the events took place. I have consulted with senior nursing experts in Canada, the US, and the UK, and the first thing that happens when suspicion is raised in this kind of case is that an outside team of medical investigators is called in, and a careful and independent investigation is made. This happens *before* the police is called in, and certainly *before* press releases are put out, alluding to activities of a serial killer.

We now know for sure that the coincidence was not due to chance because incidents were to an important extent *made to happen* during Lucia's shifts, by definition. Whether or not an event was an unexplained serious medical incident depended on whether or not Lucia was (thought to be) present. More subtly, when an incident was supposed to occur, depended on when Lucia was on duty. A full-time nurse has one of the three shifts of the day, for a number of days in a row; every incident on those days is either in or next to one of her shifts. This gives a lot of latitude for fudging the data; latitude which certainly was taken advantage of, though probably subconsciously. The investigators never specified an objective, or at least inter-subjective definition of what is an incident, nor when it is thought to occur. The court statisticians never asked to see the protocol.

An attempt to locate known incidents in shifts in a self-consistent way, without reference to who is on duty, results in several less coincidences. The original 9 incidents in a year, all in Lucia's shifts, changes on the basis of a careful investigation to 7 incidents in her shifts, 4 out of them. Since she has about one sixth of all the shifts, one would rather expect

something like 1 or 2 or 3 in her shifts, 10 or 9 or 8 out... But this is no longer that stunning "nine out of nine".

Back to paradigmas

Even if we accept the data, I will not say that a frequentist analysis is correct and a Bayesian wrong, or vice-versa. They are different, and indeed, a huge difference in numerical results, is very, very informative. Garbage in, garbage out… in this case, what you get out depends in an extraordinarily sensitive way on what you put in; it becomes clear by comparing different models and different paradigmas that the data does not tell us a very great deal at all.

Enough has been written elsewhere about *classical frequentist* versus *Bayesian* statistical inference (sometimes characterized as *objective* versus *subjective*). In my opinion neither of these paradigmas is really suitable for use "in court". I fear that neither judges nor juries are ever going to understand what is a significance level and what is a p-value (that disposes of classical statistics). I have sympathy for a so-called *empirical Bayesian analysis* but great distrust of the "true" or "fundamentalist" Bayesian use of purely subjective probabilities; especially of prior, subjective probabilities of innocence and guilt. Whose subjective probabilities? I hope that the defence will object, and their objection will be carried.

The modern alternative is the Likelihood Ratio. Actually it is an ingredient in either traditional schools' calculations; the novelty of the likelihood ratio approach is that we do not commit ourselves to what is to be done with the Likelihood Ratio. We just inform the court what it is, possibly converting numerical values (e.g., a likelihood ratio of larger than a million), into a verbal description (e.g.,the data overwhelmingly supports the one hypothesis over the other).

The likelihood ratio is simply the ratio of the probability of the observed data (the evidence), under the scenario or model adhered to by the prosecution, to the the probability of the same, observed data (evidence), under the scenario adhered to by the defence. However, this is easier said than done. No-one has yet done a decent "likelihood ratio" analysis of the Lucia data, for a very good reason: no-one has any idea of what a good probabilistic description is of events on a hospital award in a natural situation, with no killer at work, let alone in the situation supposed by the prosecution to be the case, that a certain person was trying to kill some of the patients. As I remarked above, so far everyone who has discussed the statistics of the Lucia case has taken the model of independent, uniformly random events for granted as a starting point. However, we know that 7 (=2+2+3) of the incidents concerned only 3 babies, in other words, we have sub-sequences of very similar medical emergencies occuring for the same child. We also know now that there are large fluctuations in the numbers of incidents per year over longer periods of time (a year or two before the time Lucia worked at JKZ, and again, a few years after she was removed from the scene, there were similar large numbers of incidents, to that one year whose apotheosis was a criminal investigation.

The basic empirical research into the existence and meaning of this kind of fluctuations has simply never been done, neither at those hospitals in the Hague where Lucia worked, nor in general. What we do know is that variation in the rates of events, and dependence between events, coupled with a non-uniform allocation of personnel to shifts, automatically inflates the variability in the numbers of events experienced by different nurses. Hidden

confounders leads to over-dispersion. Some nurses experience what seems like much too many, some experience what seems like much too few.

It seems that the likelihood ratio approach cannot even leave the ground. If we could get over the first difficulties, new difficulties come up; this paper is too short (and the new difficulties require more technical statistical knowledge), so I will leave it at that.

Hidden statistics

I now return to my other main theme, the persistence of statistics in the verdict of the appeal court. Here I will just summarize a number of observations made by Ton Derksen in his book and in his submission to the judicial review committee CEAS.

The court finds it significant that "all the incidents took place in a short time period and during the shifts of Lucia", when "normally there were no incidents on that ward". I think that the court here is drawing statistical conclusions (there are too many incidents, and Lucia is involved with them) from statistical data; though I have argued above that a) the data is wrong and b) the inference is incorrect, since the normal situation is one of clusters of events which then tend to "hit" one or two nurses. Of course, these observations were made by medical personel; does this make that evidence *medical*, rather than *statistical*? Actually, I don't care what we *call* it (think of Romeo's rose Juliet); the fact is that a professional statistician (I repeat: the court's experts were not professional statisticians, nor were the experts put forward by the defence) could well have persuasively argued that the cluster of events doesn't have to mean anything, nor does Lucia's presence at so many mean very much on its own, especially after we have taken care to remove *selection bias* and also taken account of the *post hoc* problem, which essentially forces us to an *empirical Bayes approach*. However, as I said before, the necessary basic *empirical* research has still never been done!

Somewhat more pernicious is the following. The court learnt from their "statisticians" that a statistical correlation does not imply causation. It learnt that the coincidence was not due to chance and that Lucia had to explain why she was repeatedly present at those incidents. The statisticians did offer a number (to be precise: four) alternative explanations of the coincidence, *by way of example.* Amazingly, deep in the "arrest" we find the judge asking Lucia, example by example, whether any of these explanations (and only these explanations) were applicable. "Mrs de B., were you a bad nurse?" "No your honour, my colleagues and superiors always had high regard for my work" [since a bad nurse would experience more incidents because the quality of her care is low, this doesn't make her a murderer], "Mrs de B., did you have to care for more difficult cases/get more difficult shifts than your colleagues?" "No your, honour, we shared the work fairly...".

My statistical colleagues and I are deeply shocked by this. The court has adopted statistical reasoning, though applying it incorrectly, and applying it to bad data; the numbers and calculations are gone, but the verbal argument of their own statistical advisors remains standing. Aernout objected.

Finally, very pernicious (in my statistician's opinion), but depending on knowledge of the underlying dossiers, is the following. For each of the deaths and critical situations considered by the court, a large number of medical experts was consulted. In every single case a majority of experts thought that the event was not unnatural, but always some experts could be found who could say that they thought that the event had some rather

strange, inexplicable features. In two cases, in which Lucia was indeed convicted for murder or attempted murder, only one expert considered the event in question inexplicable (a different expert each time). One might hope that those medical experts had strong medical arguments why these events were medically inexplicable. However, the experts said "this event appears natural, but because Lucia was present, as she was at so many other events, I believe it was an unnatural event".

This incontrovertible scientific-medical proof of inexplicable deaths or incidents is actually medical doctors' amateur statistical reasoning, based on data which we now know to be highly unreliable.

Conclusions

The fuzzy border between statistical expertise and medical or legal common sense can be used, whether accidentally or deliberately, to hide bad statistical data and bad statistical inferences away from the eyes of the world. On the other hand, the fact that in the Lucia case neither the court nor the defence worked with professional statistical experts says something negative about the visibility of the statistical profession in the Netherlands. Apparently, society at large (in this country) has little idea what statistics is about, little idea what statisticians do.

The technical problems of modelling the kind of data collected in the Lucia case are illustrative of the enormous challenges facing forensic statistics at the moment. *More research needs to be done*. I am convinced that advanced modern statistics can play an increasingly important role in criminal investigation and prosecution, and that the legal context requires the development of new paradigma's and new methodologies. We statisticians are really only just starting to tackle this job.

As criminal investigation uses more and more advanced science and technology, the legal profession is finding itself more and more challenged. This is a world-wide phenomenon. One may look for remedies in the training of lawyers, or in court procedures. More and more, forensic research is going to be multidisciplinary; more and more, scientists are going to disagree about the conclusions, and not because one is better than the other, but largely because they have different information, or are making different assumptions, or talking different languages. In science we search for differences and resolve them in open dialogue; sometimes we discover that *right now we don't know, more research needs to be done*. In the Dutch criminal court, the judges decide what information to give to one expert, ask some questions, get some answers, then repeat this procedure with another expert. From the scientific point of view, it is as if the judge is the leader of a multidisciplinary scientific investigation, where the members of his team are not allowed to communicate with one another, and one person only is responsible for the decisions concerning flow of information, and translation of information between disciplines.

I understand very well the good reasons for the existing *procedures*. However, I fear that the judge is taking over the job of a police investigation team; perhaps only attempting this because the case put by the prosecution is too weak! If the situation is such that the procedure is not likely to succeed, there is no point in going through the motions. The results are arbitrary.

References

Ton Derksen (2006) Lucia de B.: Reconstructie van een Gerechtelijke Dwaling. Veen Magazines, Diemen.

H. Küchenhoff (2008) Coin Tossing and Spinning – Useful Classroom Experiments for Teaching Statistics. pp. 417-426 in: Recent Advances in Linear Models and Related Areas.Essays in Honour of Helge Toutenburg. S. and C. Heumann (eds). Physica-Verlag HD.