Uncertainty Workshop: Everything You Always Wanted To Know About the ISO GUM But Were Afraid To Ask

R. Bentley
National Measurement Laboratory, Sydney, Australia

Abstract
The aim of this workshop was to encourage a healthy understanding of the ISO 'Guide to the Expression of Uncertainty in Measurement' (GEUM or GUM) − it is too easy to get lost in the detail of the GUM. To this end, the GUM was stripped down to its two main ingredients:

(1) basic statistical notions and concepts that underpin its recommendations and
(2) the recommended procedure for error analysis.

As a result, the purpose of the GUM may be more easily understood, various difficulties appear simpler when put into perspective, and the calculations required to get a value of uncertainty appear more straightforward. A pragmatic approach was taken with a strong emphasis on developing a healthy 'feel' for what is required and why.

The topic is fully discussed in Monograph 1 from NML, "Uncertainty in Measurement: the ISO Guide", which includes much practical advice. All participants received a copy. Further copies may be ordered from the NML web site at -

http://www.nml.csiro.au/Services/publicationsSale.htm

Introduction to part 1, The Basics
Lecture 1 covered basic statistics as it relates to measurement (chapter 1 of the NML monograph), beginning with the notion that all measurements are wrong and, therefore, that measurement is the art of handling errors. This was followed by a discussion of error, uncertainty, confidence limits and the need for probability. A histogram of DVM data illustrated an error distribution and another showed that systematics have distributions. Finally, how to cope with random errors (reduce by forming an average, etc.) and systematics (be aware, reduce, calibrate, etc) was discussed.

Discussion (part 1, The Basics)
There was no discussion following part 1.

Introduction to part 2, Around the Edges
Lecture 2 covered just those concepts and facts needed to understand the ISO Guide (chapter 2). Some definitions are needed (measurand, type A & B, standard and expanded uncertainties, coverage factor and degrees of freedom, etc). The concept of degrees of freedom was discussed – it expresses the quality of the SD as an estimate of sigma and, thus, can be infinite in only a few cases – and a pragmatic approach to choosing its value was suggested. Modelling a measurement and the need for sensitivity coefficients was also dealt with.

Discussion (part 2, Around the Edges)
David Burns – I see that the uncertainties have no sign. Don't you normally write them as ±?

Robin Bentley – I know. I didn't say the subject was easy did I? Uncertainties …. OK, I was going to mention the topic later, but I'll do it now because it's appropriate. Most people regard an uncertainty as being a region within which you're likely to find the number, the error or the measurement result. So, they expect to see ±x, and quite rightly. But you see, in the ISO Guide, standard deviations are used and we have terms like standard uncertainty, which is a standard deviation. And the standard deviation has no sign nor therefore, does the standard uncertainty and the expanded uncertainty. So, it would be wrong for you to state in a report that "the expanded uncertainty was ±6". It's not, it is 6! On the other hand, if you said that the expanded uncertainty was 6, the average person would take this to mean ±3! So it's ambiguous!

What I suggest, then, is that we always use ± because the reader would otherwise misinterpret, but we don't use statistical terms or the special adjectives of the ISO guide. Use normal English: "The uncertainty of the calibration was ±6" rather than "the expanded uncertainty was 6". It's a small point. So, what I'm saying is, don't use the adjectives of the ISO guide in any report for items with a '±', because those adjectives refer to things that have no sign, and if you take the sign off, people may well misinterpret.

Nkem Anele – You spoke about the random and the systematic error or the type A and B, how do you treat the errors if you cannot combine them, whether random error, systematic error or type A and B.

Robin Bentley – I didn't quite understand that, I'm sorry.

Nkem Anele – Resolving type A and type B, how do you combine random error and systematic error.

Robin Bentley – What I said was that type A and type B uncertainties should be combined, they shouldn't be treated separately, as they may have been in the past. They are calculated differently, but once you have obtained the standard deviations, you combine them. You don't differentiate. Once you have determined the standard deviation, whether it's via a type A determination or a type B, once you've done that, once you've got standard uncertainties, you treat them all the same. The way you calculate them, the way you discuss them and consider them is different, but don't treat them separately.

Introduction to part 3, At Last The ISO Guide

Lecture 3 described the calculation process given in the Guide, as seven steps beginning with forming the model and ending with the statement of results, and gave an example calculation (chapter 3). The example was extended to demonstrate that the number of degrees of freedom assigned to an uncertainty is not critical. As long as its value is in the right ball-park all is well!

Discussion (part 3, At Last the ISO Guide)

Kent Gregory – Just a question about Excel. If you don't use Excel or a spreadsheet, what do you use? Do you do it longhand or (use a) calculator?

Robin Bentley – Well, I don't think about this much any more. But, generally, in our lab we have a set number of calibrations or types of measurement to deal with. And we have work instructions on how each one is done. And the calculation of uncertainty is made once for each – once it's done, that's the end of it. So for me it doesn't matter much that a manual calculation had been used and took five minutes. It's done only once every couple of years ... If you have to do it a lot, then, sure, use a spreadsheet like Excel.

However, don't take it on face value, check it out first. There are a number of articles from NPL, for example, that have highlighted some errors in the use of such spreadsheets, much the same way as occurs with calculators – or programs such as PV Wave or Origin. Behind the scenes there is internal rounding, for example. ... It depends on how you make your calculation – you can be subtracting two large numbers, for example, and rounding at the wrong time may lead to a zero result. You might need double precision. So, just check out the program, before you bother using it.

Quite often, of course, when it does make a mistake, it's a ridiculous result – so you can tell. But a lot of novices can't tell if a value is ridiculous. When I see a number of degrees of freedom ν_{eff} of more than 100, I suspect the person has made a mistake. A lot of people end up with large values because they have assigned infinite degrees of freedom or 10^{14} to a variety of errors that they've overestimated. They feel confident that they've got the thing trapped in there somewhere, so because it's in there, "I'm going to assign infinite degrees of freedom to it". But that's breaking away from the principle behind the Guide, of statistics. Degrees of freedom represent the quality of the estimate. Not whether you've got it or not, but whether it's the same as σ.

Nkem Anele – In Rounding off, do you have to leave it to the end?

Robin Bentley – Rounding, yes ... In the next talk I shall discuss this, with data fitting, not because you might be interested in data fitting, but because there's a trap to it; and a simple way of looking at correlations. Though, it doesn't look all that simple – but you can look at correlation effects in a particular way that may help. Also, I shall discuss the complication of symbols, and suggest an easy way out there. I like easy things. So that's what I'm going to talk about next.

Michael Kramer – In one of your first slides in the last presentation, you said that the model could be incorrect. That is a difficult issue because you don't know how wrong your model is. That ... addresses whole issues that you haven't understood something. If you know that your model is wrong, then you choose a better model (in) which that is right. But what do you do with a wrong model?

Robin Bentley – It's starting to become more interesting isn't it? It's the same with all your error estimates. You don't know the value of any error do you? You have to make an estimate ... You make use of your experience, other peoples' experience. If you can't get a guide from that, you start tweaking things and seeing what happens. You can do experiments to see whether the model is good enough. It's just the same as the other error estimates you make. You have to do the best you can. And sometimes you just toss a coin and make a guesstimate -- based on your experience. If you're not confident, ask someone who has more experience. It boils down to doing the best you can with the data you have.

One way out of it is to be very careful of how you specify your result. You could say, "My measurement according to that model is this" Or you can say, "the output of that equation is this ± that". You're not saying the model is right. You're just saying, "The calibration of this instrument *today* - not tomorrow, just today - is so and so". It is more easily determined. But the client wants a calibration result for tomorrow. You may not know enough about the drift to predict that, which is what you say when he complains about it. This is where the model comes in – that's what specifies what you mean by your measurement. So you can avoid some very hard issues by saying, “the output of that equation is 76±1”. And then it’s up to the reader to decide if that equation is applicable to him. But usually, you can put numbers in – from your experience – you’re in the best position to do that.

Introduction to part 4, Finishing Touches
Lecture 4 finished the workshop by addressing a number of loose ends. Important topics, such as possible confusion in the use of symbols, with a suggested solution (chapter 2), a pragmatic approach to correlation effects (chapter 2) and, from chapter 4, the rounding process and a problem associated with curve fitting.

Discussion (part 4, Finishing Touches)
John Baldas – I’ve got a question about least squares fitting. You minimise the squares of the deviations in the y. What happens if you minimise the squares in the deviations in the perpendicular distances of the points to the best fit straight line. Do you get the same result or not?

Robin Bentley – I didn’t quite get that I’m sorry. My hearing must be off. Could you repeat?

John Baldas – In the normal least squares, you minimise the difference in the y parameters - of your y point and the ... line of best fit. Now, if you did it another way, you minimise the perpendicular distance between your point and the line of best fit, would you get the same answer, because what you’re saying is, that in drawing the line of best fit, if you do it by eye, you tend to minimise the perpendicular distances, if you do it mathematically, you tend to minimise the ... squares of the vertical distances. So, are the two equivalent?

Robin Bentley – Probably not. Mathematically the vertical one’s far easier to consider.

John Baldas – Yes. I would have thought that the perpendicular distance would be ... the one that you’d choose, and if you did that with your sine curve ... when you twisted everything around and made it ... a parallel line ... you gave the impression that you were actually minimising the perpendicular distances, which you weren’t ... sum of the squares ...

Robin Bentley – In that particular case, the sine curve would not have been visible on the initial plot. The sine curve is a factor of maybe 50 smaller than anything you can see on that first plot. Therefore, the result of a least squares fitting program would not have revealed that variation on the plot.

John Baldas – I assumed when you do some measurements ... if you have a system where you’re actually smoothing lines and if you do have a sine curve, and ... the number of points is too low, you can end up with exactly the same problem.

Robin Bentley – I don’t want to answer some of these questions if they’re too far removed from a context that I’m familiar with ... I’m partly trying to escape from answering your question ... but also it’s rather difficult to answer a question out of context. In any case, as far as data fitting goes, the fact that you do it by eye doesn’t necessarily mean it’s better. When doing it mathematically, it’s so much easier to be confident in the answer, because everything is understood mathematically – you can work out standard deviations, apply chi squares, etc ….. Sure, you might think intuitively, or whatever, that by taking the perpendicular you'll get a better result, because it seems to resemble doing it by eye, but is doing it by eye better? I don't know that. I'd rather follow what the maths suggests because that's clear.

In any case, the point that I was making wasn't how close you can get to the data with your fit, or how good that fit is, but what you do when you get it! You should plot the differences and it's what you do with these differences is what I'm talking about. If there are less than 10 or so data, you can't answer the question, "How much of the scatter about the fit is random and how much of it's systematic?" It is not whether I can improve those differences a fraction by fitting a different way, but what do you do with the differences when you look at them? I'm suggesting that with little data you can't rely on the confidence limits given by the program, however you do it. It is more important to look at the differences from the fit. All you can say is that there is a possible systematic of plus and minus half the range, or, if this is too large, get more data. That's what I'm saying.

David Burns – Doesn't this come down to what you were saying earlier about the model, because when you put a linear fit through you're saying that's my model, I've got a linear model. And according to that model you can evaluate an uncertainty, and on top of that you may want to put an uncertainty on your choice of model.

Robin Bentley That's right, ... but in this particular example, and this is what happens in many cases when you calibrate, the model is the equation you use for each one of those data points. Each one essentially has a separate 'model'. It's the same equation, but calculated at different values. And having got your uncertainties, or your corrections, you plot them to see if you've missed something. And you note a significant scatter and what do you do with it? A lot of people would take the least squares fitting confidence limits, or standard deviations, and just feed them in as their fitting uncertainty. What I'm saying is that that's irrelevant. It's the wrong number. You should put this bigger number in.

Malcolm McEwen – That situation there, was saying that if your individual uncertainties on each point are much smaller than your fitting then the model's wrong and you can't ...

Robin Bentley – The model may not be wrong, it's just that there are uncertainty components you haven't considered. And that's where it's helpful ... I'm not addressing the subject of fitting because fitting is best done with a large number of data, if you want to find the function that applies. As I state in the book, we're only talking about the case that happens quite often, say in calibrating equipment, where it's too expensive to get a hundred calibration points on a range of one instrument. The customer wouldn't pay for that. So you're limited to, say, half a dozen data points. Or, it might take too long to do other than five points. And what do you do with those five numbers when you get them? Sometimes, the plotting of them does no more than highlight an error.

Once, I went to a company that had a reel of calibrated wire – thermocouple wire – and they had attached a hand written tag with calibration data. This is bad, because of possible transcription errors. Each reel had corrections at four different temperatures – the temperatures they used the wire at. As I always mentally plot things without thinking about it, I looked at those four numbers and realised that one of them was wrong – it would have stood out on a plot. They hadn't noticed because they were looking at the numbers themselves. On the other hand, if they'd plotted the data they would have seen that one was way out – the departure was unrealistic, and they ought then to have gone back and checked it out. So, even if for no other reason, you should always plot.

In any case, back to the point you were raising, if you can predict the scatter, as I said earlier, by looking at your list of uncertainties, in the analysis table, and pick only those bits of them that affect an hour-to-hour fluctuation, you can by eye predict roughly the kind of scatter you'd expect on plotting all the corrections. So, if the ± 2 units (on the plot) can be predicted from that, then you've covered it. However, you may well find as often happens, there is more scatter than is predicted from the other components. Then, there is a systematic or two at play, varying with the instrument – it might be your measurement technique or something's come into it that you're not aware of, something that's affecting all those results. They ought to be considered. In this case, you're not trying to determine the fitting 'accuracy' or – what shall we call it? – the interpolation error. In other words, don't just use a value given by the least squares fitting program.

Lew Kotler – Just one question, (when) you talked about the extended uncertainties, you gave it as a value rather than a percentage. Do you recommend using a value rather than percent – rather saying it's 10 plus or minus 0.6 or something, ... rather than saying 10 plus or minus 0.4% or whatever value it comes to.

Robin Bentley – I 'never' say percentage ... I don't like percentages – you've only got to listen to politicians or anybody else in the media. I often wonder whether they're talking about percentage of a percentage. For example, if the unemployment rate is 6% and has gone up by 0.4%, what does it mean: 0.4% of 6%, or has it gone from 6 to 6.4%? Percentage of what, is the thing. So, that's why I think metrologists should always express an uncertainty in the units of the measurement.

Lew Kotler – Not percentage?

Robin Bentley – Not percentage.

Malcolm McEwen – But then you have to put individual uncertainties on every point.

Robin Bentley – That's right. But just as I gave you in that example, it's possible, if the uncertainties can be plotted, and you can express them as $a+ bt$, or whatever, say a linear function, then you can express the uncertainty as not so much a percentage full stop, but as a factor. Suppose we have ± (5 g + 0.1% of the mass) ... it needn't be called %, it is instead better expressed as '0.001 M'. So, here we have an option that gives you the value of uncertainty involving a percentage, but where it's clear what you mean by the percentage. In any case, I wouldn't say, "I've measured 500 g with an uncertainty of 0.1 %". I would rather keep it in the same units. All that the Guide says about it, I think, is that you express the uncertainty in the same units as you express the measurand. And you express the units in both statements, so you put: 6 g ± 0.1 g.

David Webb – Surely you have to look at what ... the calibration value you are providing will be used for. If it's to be used in a multiplicative fashion, then the user will have to ...

Robin Bentley – Yes ... I'm just wary of percentages because they're ambiguous, often. There is always the question: percentage of what? Although, if you express what the "of" is, that's fine.

David Burns – I think the Consultative Committee on Units has addressed this problem of percentages. And they discourage the use of percent, but you can use relative standard uncertainty. But not express it as percent just because confusion can arise in many cases with percents.

Kent Gregory – Just a quick question about the final reporting of measurement uncertainty. You mentioned two examples. A digital voltmeter where it's been calibrated and you're using it in a particular fashion to minimise the uncertainty so you're keeping it on the same range, you haven't got autorange on, and you're doing say a ratio of two different voltage measurements or something, so really the uncertainty with your measurements is overestimated if you just use the uncertainty stated on the calibration report because that's taking into account things like drift and range selection - differences in ranges and so on, so ... where do you get the information short of ringing up and bugging the people that did the calibration in the first place ...

Robin Bentley – Well, I think it's a case of when you make the first contact with the calibration lab, it's important to specify exactly want you want. Let them know, and they can design the calibration to suit. It's a communication issue ... if I rang a calibration lab and said I had a thermocouple to calibrate and they didn't ask me any questions, I wouldn't send it – because they don't know enough to ask the right questions. There are too many issues with thermocouples. The devices are two dimensional, in the sense that if you move them their calibration changes – their calibration depends on the temperature profile. So I wouldn't trust a calibration lab that doesn't ask the right questions.