Misconceptions about Error

COMMUNICATION

John Garratt, Andrew Horn, Jane Tomlinson,

Department of Chemistry, University of York, Heslington, York, YO10 5DD Telephone: 01904 432551 e-mail: cjg2@york.ac.uk

First year students have been introduced to Least Mean Squares Linear Regression as part of a laboratory exercise in which they determined $\Delta H^{\ominus}$ from measurements of an equilibrium constant at different temperatures. We took the opportunity to obtain feedback concerning student understanding of the value of using an objective (statistical) method for fitting a line to experimental data and of the meaning of 95% confidence limits. 65 students have provided feedback by writing short comments as part of their laboratory report. Our analysis of these responses indicates that a majority of the students believe that the use of LMSLR increases the accuracy of (or reduces the error in) the calculated results, and that a large proportion hold confused views of the meaning of confidence limits. We conclude that these misconceptions are illustrative of a broader range of misconceptions about the origins and consequences of experimental error, and that these are significant barriers to learning.

Introduction

Goedhart and Verdonk[1] report that their first year students *"experienced large difficulties in performing error calculations"*. From their description of the difficulties, it appears that they expected their students to have gained from lectures a rather sophisticated understanding of statistical methods used in error analysis. However, their research into students' interpretations of statistical concepts indicated that their expectations were misplaced. They concluded that *"students interpret errors in a personal context: they think they are responsible themselves for measurement errors. This interpretation was not changed after the lectures in statistics, even if attention was explicitly drawn upon the meaning of error in statistics"*.

Our own experience with giving lectures on statistical procedures was sufficiently discouraging that we tried an empirical approach. We wrote a computer program which allows students to investigate for themselves some key concepts of common statistical procedures[2]. An underlying assumption behind our computer-based package is that students have a general appreciation that random error is a feature of experimental measurement, rather than being what Goedhart and Verdonk refer to as 'personal errors'. Our package had a mixed reception from second year students[3], and this may have been partly because we had not taken proper account of the students' misconceptions about error and error analysis. Therefore, when we devised a new strategy for introducing our first year students to the potential value of Least Mean Squares Linear Regression (LMSLR), we decided to give them the opportunity to reveal some of their misconceptions most relevant to the exercise in question.

We report here our initial conclusions, which have stimulated us to undertake a more detailed study.

Methods

The laboratory exercise which we intended to use to introduce linear regression involved measuring the equilibrium constant for the dissociation of ammonium carbamate at different temperatures, and calculating $\Delta H^{\ominus}$ for the reaction from the slope of a graph of lnK against 1/T. We assumed that all students would appreciate that, where a linear relationship between two variables is expected to exist, real data will not fall on a straight line because of the existence of experimental error. We further assumed that (most) students would have some experience of choosing a line to fit real data, though we expected that different students would have different criteria for judging the 'best' fit. We did not expect (many) students to be familiar with the concept of confidence limits, or with statistical (objective) procedures for drawing straight lines.

We took the view that, during the students' first term, it was inappropriate and unnecessary to deal with the mathematical basis of statistical theory and practice. We wished the students to appreciate the following key points:

- Fitting a line 'by eye' (i.e. subjectively) is not satisfactory because the criteria of 'best fit' cannot be described, and therefore any subjectively fitted line cannot be reproduced.
- There is a correct value for $\Delta H^{\ominus}$ for the reaction concerned, but experimental error leads to lack of confidence that the 'best' value for the slope gives this correct value; it follows that it is useful to quote a *range* of values that is likely to include the correct value. (We are aware that statistical procedures for fitting lines to data cannot take account of systematic errors).
- Least Mean Squares Linear Regression (LMSLR) is a generally accepted procedure for fitting straight lines to data; computer programs can use the criteria of LMSLR to calculate the 'best fitting' slopes and intercepts together with 95% confidence limits to these values.
- Two criteria are used to fit lines by LMSLR. One is that the line passes through the point x, y. The other is that the slope is determined by minimising the sum of squares of all the differences between observed and calculated values of y. Whether these criteria are or are not appropriate to the data is a matter of judgement.

- 95% confidence limits can be decreased by reducing experimental error or increasing the amount of data collected.

We decided that the best way to help students to understand these general principles was through the use of computer software which we developed for this purpose. This offered the following specific benefits:

- The computer allowed unique opportunities for the active involvement of the student in the learning process.
- The program could easily provide students with different sets of data to examine.
- The computer would force us to rely heavily on a visual rather than a verbal approach to explanation, and this was more likely to be more successful with this particular topic.

An important constraint was that, in order not to overload the students with extra work, it must be possible for them to complete the laboratory work and the computer-based exercise within the normal laboratory day of 5 hours, and this meant that they should spend no more than 20 minutes at the computer. We will not describe the software in detail since it is being substantially revised in the light of our subsequent evaluation of students' understanding of errors. However, our analysis of student feedback after they had used our software revealed some of their misconceptions about this subject.

We attempted to draw attention to the lessons we wished the students to learn by including the following two statements in the laboratory handbook. Both written statements were reinforced verbally in the pre-exercise briefing.

The computer program will help you to think about

- *The criteria used to calculate the line of best fit to your data as defined by the least mean squares linear regression (LMSLR) technique;*
- *The reasons for using statistical methods for estimating error and their effects upon your calculation of $\Delta H^{\ominus}$;*
- *The reason for quoting confidence limits.*

As part of your write-up, you should produce a paragraph summarising

- *The reasons for drawing a straight line through data using an objective rather than a subjective method.*
- *The meanings of 95% confidence limits on the values for the slope and intercept.*
- *The advisability or otherwise of using the LMSLR procedure as a routine method for determining the best fitting line to data.*

You should include a comment on the thought you have given previously to these aspects of error analysis and how useful you found the computer program in helping you to develop your thoughts.

We had two reasons for choosing the free response format for obtaining feedback. The first was that it takes more reflection to write a free response than it does to complete a fixed-response questionnaire, and we judged that this reflection would improve the student learning experience[4]. The second was that we expected the free-response format to reveal ingrained misconceptions which might not have become apparent form a fixed-response questionnaire. We recognised that free-response questions do not easily provide quantitative data. However they go some way towards formalising both the observation and the noting of critical incidents which Goodwin recommends as an alternative to the full-scale recording and analysis of interactions in a classroom[5]. The analysis is inevitably time-consuming, but we felt that the potential advantages outweighed the disadvantages.

Results

We received photocopies of 67 scripts which included the student responses to the task of writing a paragraph about errors. Of these, we discarded two because the students had badly misinterpreted the task. Not all of the remaining 65 responded directly to all three points they were asked to cover. The responses are summarised below.

Reasons for drawing a straight line using objective method.

Forty two of the students showed that they appreciated two related disadvantages of using a subjective method: it is not possible to guarantee that the line can be reproduced, and the criteria used to draw a subjective line cannot be described precisely. Some of these explanations were not particularly well expressed (for example, two merely stated that an objective method made it easier for others to interpret the data). The underlying point that the objective method uses *defined* criteria to fit the line to the data was made explicitly by thirteen students; twelve of these were included in the forty two. Of these thirteen, three included both of the criteria used by LMSLR and two included one of them (even though they were not specifically asked to do this). In addition to these five students, a further ten gave both criteria and ten more gave one of them, but these twenty students did not comment on the advantages of being able to define the criteria. This suggests that students find it easier to remember the specifics of the criteria than to explain why it is important to have defined criteria.

Eleven students commented that the least means squares method 'assumes' three features in the data:

- that the error in x is zero (or negligible);
- that the errors in y have a normal distribution;
- that there is no systematic error.

Only three of these students appeared to be aware that these three features are not *assumptions* made in carrying out the procedure, but that they define the conditions within which the procedure is justified – an important distinction for them to learn. Only four students in total made the point that LMSLR may not always be the most appropriate way of defining the line of best fit. This again indicates that it is easier to absorb specific information such as the criteria described above than to understand its meaning.

The most common misconception about the advantages of using an objective method, in this case LMSLR, is that it somehow improves the accuracy of the best fit line, or that it removes experimental errors. A typical comment is that the best fit line "*will be far more accurate than if found by hand*

and eye", or that "*The line being drawn by this technique will therefore lead to less experimental error...*". In all, 33 students made comments of this type. This suggests that, as concluded by Goedhart and Verdonk[1], our students have a very confused understanding of the nature and origin of errors. It also suggests that first year students have not appreciated that error analysis involves more than the mechanical application of statistical procedures but requires judgements about the most appropriate methods to use.

Seven students raised the issue of anomalous results in ways which indicated that they have no clear understanding either of what these are or of the effect they have on results. Illustrative comments are

"It has the disadvantage of using every piece of data";

"The problem with LMSLR is that anomalous results could severely undermine the accuracy of the straight line";

"It also ignores misalliance results".

These, and other comments which imply that errors occur in selected data only (rather than randomly in all data), are consistent with Goedhart and Verdonk's suggestion that students regard errors as personal.

Five students made reference to the problem that the presence of error makes the theoretically linear relationship hard to see. They appear to be chasing a correct result (known in advance) and consequently feel that any data which undermines this 'rightness' is an anomaly and should be discarded. This again indicates a confused view of the nature of experimental error and how it should be treated.

The meaning of 95% confidence limits

Nineteen students made statements which showed that they understood that, in this context, confidence limits define a range of possible lines which provide an acceptable fit to the data, and that there is a 95% probability that this range includes the correct line; a number of them express this as a 1 in 20 chance that the range will not include the correct line – a perfectly acceptable equivalent to 95% probability of inclusion. A further five students described the confidence limits as providing a range of possible best fit lines but made no comment on the probability of this range including the correct value. This may indicate an understanding of a point made by the software that the range can be calculated to give any reasonable level of confidence. We would not expect at this stage that students would show deep understanding that 'correct' has no clear meaning, depending as it does on the assumption that LMSLR is the appropriate method for analysing the data.

Nine students made no significant mention of confidence limits, and thirteen made comments which were so vague or confused as to confound any attempt at interpretation.

The remaining nineteen students showed various misunderstanding of the 95% confidence limits. Seven appeared to believe that the confidence limits provide a range within which there is a 95% chance that all data points will lie; they did not appear to associate confidence limits with confidence in the slope and intercept. Six thought that the range given by the confidence limits means that the correct value lies within +/- 5% of the best fitting value and six others refer to being 95% certain that the calculated result is correct.

Typical statements which illustrate the kind of confusion which arises are:

"…the values could be + or -5% of what was recorded";

"95% confidence limits ….leads to less experimental error";

"95% certain that you have drawn the straight line in the correct place";

"…reflect the ranges in which 95% of the experimental data will fall";

"The confidence levels shows how close the line fits the position of the points";

"…tell us the possibility of a real point being outside the set area".

Previous Experience

Seventeen students made general comments about their previous experience of error analysis. The great majority of these indicated that the respondent had not previously encountered objective methods either for fitting lines to data or for calculating confidence limits. It is likely that this is true of a similar proportion of all students and is not restricted to those who commented. A few of these comments revealed other features of students background beliefs which may be more widespread. These include

"I realised that drawing a best fit line does have errors, yet I had not realised how much of an error it was";

"I have used LMSLR in maths but never applied it to real scientific data";

"Previously I have used the correlation coefficient as a way of calculating the straightness of a set of points";

"I had previously been led to believe that by drawing an estimated line of best fit, I somehow eliminated the error".

Discussion

The student responses to our computer-based introduction to LMSLR demonstrate widespread misunderstandings about the factors affecting accuracy in experimentally determined parameters and about the meaning of confidence limits. These misconceptions almost certainly reflect a wider range of misconceptions about the nature and origin of experimental error which our exercise and the feedback from it could not have brought to light. We are currently planning a more broadly based survey with a view to establishing the extent of these misconceptions. However these present results are significant in that the responses were obtained from students in their first term at university, almost all of whom had completed their A levels in the previous summer. Their misunderstandings of the treatment of error cannot have been significantly influenced by their university experience and we therefore conclude that they are typical of first year chemistry students throughout the country.

The widely held belief that LMSLR increases the accuracy of (or decreases the error in) the final result almost certainly arises from the conventional use of the phrase 'line of best fit'. For a student drilled to appreciate the importance of accuracy, it will seem natural to associate 'best' with 'most accurate'. In fact, given the limited amount of data collected, it is unlikely that the best fitting value will be the correct value (even if there is no systematic error). This is taken into account by the Confidence Limits, which are also misunderstood by many students. Both these points require a sophisticated understanding of statistical procedures which it takes experience to develop. A further conceptual sophistication is the point that a (more or less) subjective decision is taken to define the criteria of 'best fit'. For example, in our view it is unlikely that the data collected in this laboratory exercise actually comply with the criteria used to calculate the LMSLR. In this connection, many of the students themselves commented that a major source of experimental error was the difficulty of measuring the temperature of the apparatus, but none pointed out that this was contrary to the assumption that there is no error in the vale of 1/T. However we felt justified in using the opportunity provided by this exercise to introduce this important procedure, and we judge that it is of greater value to overlook the technical inappropriateness of the procedure than to introduce an additional complication.

In preparing the specification for our software we made no attempt to introduce the concept of accuracy. However, it has become clear that student misconceptions of the specialist meaning of this word are a serious barrier to an effective understanding of the virtues and limitations of LMSLR.

In contrast to the lack of coverage of the concept of accuracy, our software was designed to help students to understand the concept of confidence limits. Nevertheless, it is clear from the student responses that misconceptions about confidence limits persist in students who have used our software. Almost certainly these misconceptions arise from confusion of the use of 'confidence limits' in statistical analysis with the use of 'confidence' in everyday language. Here, 'percentage confidence' (or 'certainty') usually refers to confidence in knowing or having a correct answer (as in "I am 95% confident that I turned the cooker off"). The percentage is not normally associated with a range. Thus, everyday language is unlikely to produce a sentence such as "if I toss a penny 100 times I am 95% confident that the number of heads will be between 46 and 54". Even more alien to everyday language is the apparent paradox that the average number of heads tossed is likely to get closer to 50% as the total number of tosses increases, but it gets increasingly unlikely that the value will be exactly 50% (there is a good chance that two tosses will result in one head and one tail, but only a very small chance that 1000 tosses will result in 500 heads). We conclude that it is a particularly sophisticated scientific concept that it is normally more useful to define a range of numerical values which will (most likely) include the correct value than it is to attempt to define exactly the correct value. The concept is reflected in the caption of a cartoon which reads *"Do you want a 100% guarantee it's 99% pure or a 99% guarantee it's 100% pure?"*[6]. Students are likely to find the concept of confidence range particularly difficult since they are consistently presented with numerical values in which no errors are admitted (tables of Relative Atomic Mass, Standard Redox Potential, pK values, etc). In this context it is not surprising that spending 20 min using a simple computer program is insufficient to correct ingrained misconceptions of the scientific meaning of words which are used more loosely in everyday speech.

The constructivist view of learning holds that "*knowledge is constructed on the mind of the learner*" by integrating new knowledge with existing concepts[7,8]. It follows that effective learning is unlikely in a mind which is already full of misconceptions, unless proper account is taken of these misconceptions. We have argued here that first year chemistry students hold many misconceptions of the use and meaning of language used to describe and apply error analysis. We suggest that teachers disregard at their peril the first of Johnstone's 'Ten Educational Commandments' that "*what you learn is controlled by what we already know and understand*"[9]. We are also encouraged by the evidence that some of the objectives we set ourselves have been met by our software, and we conclude that it will be worthwhile to extend and develop this approach in order to overcome fundamental misconceptions about error.

Acknowledgements

One of us (JLT) was supported by the ESRC under grant number RO22 25 0121

References

1. Goedhart MJ and Verdonk AH 1991 The development of statistical concepts in a design-orientated laboratory course in scientific measuring *J. Chem. Ed.* 68 1005–1009.
2. Garratt J and Smith S 1994 Empirical illustration of statistical procedures *Biochem. Ed.* 22 15
3. Garratt J, Clow D, Hodgson A and Tomlinson J 1999 Computer simulations and chemical education *Chem. Ed. Rev.* 14 51–81
4. Tomlinson J 1998 Reflecting on learning *U. Chem. Ed.* 2 35
5. Goodwin A 2000 Some thoughts following "Crossing the Borders" *U. Chem. Ed.* 4 37–38
6. Garratt J, Overton T and Threlfall T 1999 *A question of chemistry* (Longmans, Harlow)
7. Bodner GM 1986 Constructivism: a theory of knowledge *J. Chem. Ed.* 63 873–878
8. Taber K 2000 A constructivist approach to university chemistry teaching *U. Chem. Ed.* 4 63–72
9. Johnstone AH 1997 Chemistry teaching – science or alchemy *J. Chem. Ed.* 74 262–268