EXPERIENTIAL LEARNING IN STATISTICS THROUGH THE COMPUTER

Walter A. Smith, The University of Tulsa

Statistics in the classroom is not a spectator sport. The participants must participate -- and therein lies the problem. Almost any classroom example of the use and computation of any statistical concept regardless of the small size of the data set and the small numbers making up the data set soon becomes a very large number of very large numbers. The amount of computational tedium associated with student solution of assignments often becomes overwhelming, not challenging. The mathematical techniques are elementary. Their application is boring at best to the accomplished student and mystifying to the unaccomplished. The end result is that, except in rare instances, we fail to reach even a majority of students. In a recent vernacular: we not only fail to turn them on, we turn them off.

Any salesman knows that he makes progress towards a sale if he can get the prospect to "feel the goods." The prospect must DO something. Much of what a teacher does is to sell an idea. Some are relatively easy to sell. Some are difficult. The difference is often related to the amount of willing classroom participation and how engaging the outside assignments are. Some subjects lend themselves quite readily to the "gentlemen's debate." Statistics does not. Nor are the usual arithmetic type problems engaging. So we must entice the student into DOING something that will be engaging and more importantly, challenging. We must reduce the tedium of the arithmetic and make the assignments timely and meaningful. The computer can offer a powerful assist in reaching these ends.

THE CHALLENGE

The challenge Is two-fold. Number one is to the teacher; number two is to the student.

Number one

Since teacher evaluations have become the current fad in academia (for good or for evil), the teaching profession has become more aware of the student. It is likely some of the "duty" to the discipline is becoming "duty" to the student. If this is the case the teacher must seek for ways to involve the student. He can no longer simply lay it out. He must enlist the student in the cause of the discipline. The challenge to the teacher is to make his class a personal encounter for the student.

A change is taking place on the other side of the lecturn, too. The social revolution that has been taking place has brought a new kind of student to the campus. He is a more aware person. He is more inquisitive. He is demanding. He is no

longer in school because it is the socially acceptable thing to do. On the contrary, he is often going counter to the mode of his peer group. His very presence adds to the challenge to the teacher. His attitude is: Here I am, prove to me I am not wasting my time. The challenge to the teacher is clear.

Number two

The challenge on the other side of the lecturn is not so clear. The student is in the statistics class under duress. He is not a willing participant. Whereas the challenge to the teacher is obvious and self imposed, the challenge to the student is not clear and must be imposed by the teacher. But it is not likely to be accepted unless it is a dare that is unique. There is no peer group pressure to accept the challenge. On the contrary, the opposite is often true. The student must be challenged to DO something. As a result of DOING the student must receive some personal satisfaction that is unique to him. It must not be the promise of a reward in the future, or the fear of some penalty In the future. It is now.

A RECENT EXPERIMENT

In order to Induce the student to accept the challenge, four elements in the approach must be present.

1.   The learning experience must be unique.

2.   The matter under investigation must be timely.

3.   The student must not be overwhelmed with computational tedium.

4.   The response time with respect to either acceptance or rejection of his efforts must be as short as possible.

These elements were later paraphrased into the objective of the system that was to be developed.

The Vehicle for Learning

Statistical estimation of the arithmetic mean is the first contact with inference that the student makes. It is a fairly straight forward application of his recent review of probability theory and his introduction to sampling distributions. It is essential that he gain a solid understanding of the process before he moves on to further application of the inferential process. Accordingly, this concept was selected for the first challenge.

So that all the elements of the challenge could be included, an inter-active batch mode computer program was written (Statistical Estimation of the Arithmetic Mean, SEAN). It is admitted that response time could have been materially shortened if the inter-

active computer terminal had been used instead of the batch-mode. However, the nature of the exercises to be computer-generated would require much more time at the terminal than was deemed economical of its use.

The Computer Application

The concepts associated with statistical estimation are introduced in the classroom lecture along with instructions with respect to accessing the computer program. The student inputs the initial requirement consisting mainly of the details of program access and student identification. The computer response is a personal letter from a "client" to the student in which the student is named the principle investigator of a problem encountered by an organization in which he has an interest. His interest was identified in his initial input. This letter also outlines the procedure to be followed in conducting the investigation. It is, in fact, a paraphrase of what might be found in almost any introductory text. There is a difference, however. The student must read it in order to assuage his curiosity that has been aroused, even if only slightly. Moreover, the continuation of the exercise is conditional upon his becoming familiar with it. In addition, the letter sets forth certain requirements generated by the client for whom the investigation is being conducted. These are pre-set values for the Level of Risk, fwso~ , and the Tolerance, E. The instructions also indicate to him that the variance is known and he is provided with its numerical value. (See Chart 1, Flow Chart for SEAN). He then proceeds, according to instructions, to determine the minimum sample size suitable for meeting the requirements. This data point is returned to the computer whereupon it generates a unique sample of the size specified. Sums of the variable, its squares, and deviations are also provided so that the student may calculate the mean, standard deviation, and standard error of the mean without the usual tedium.

It is at this stage of the exercise that uniqueness is provided. A random number generator furnishes each student with a different set of data. The student ID number, previously entered into the system is used as an index number for a DATA statement which determines the several control elements available in the system. An internal security system prevents any student or combination of students from accessing the same data. The required calculations are made by the student and his card deck is submitted to the computer center, where his arithmetic is checked. If his solutions are correct, a pictorial representation of his findings is presented (Chart 2) and he is instructed to make the test to determine if the tolerance he has calculated is less than or equal to the tolerance initially specified by the client. In this first exercise the calculated tolerance will be less than the initially specified tolerance and he proceeds as instructed by a Memo from the computer to interpret his findings and to prepare a written report to be submitted in class on a specified date.

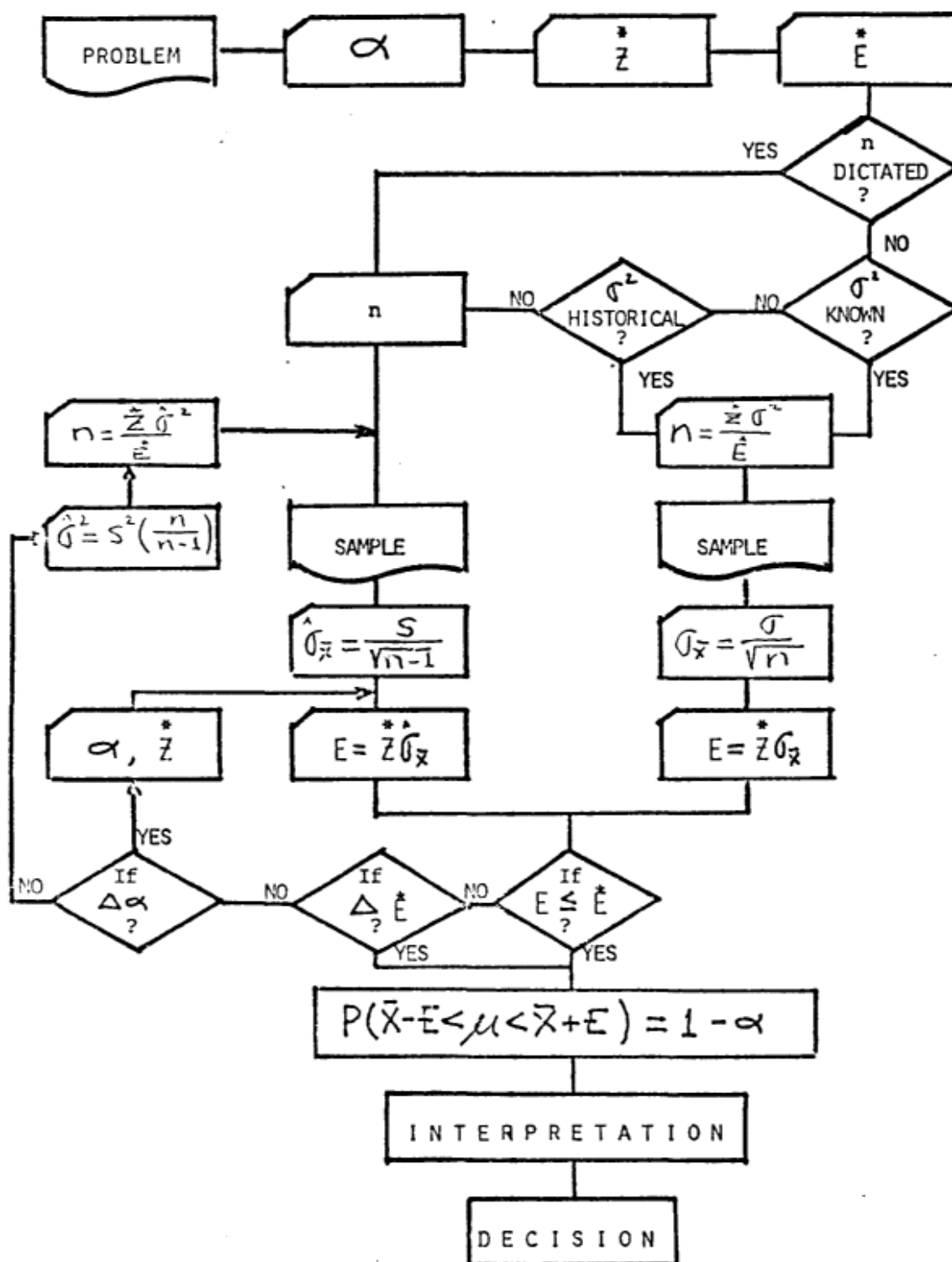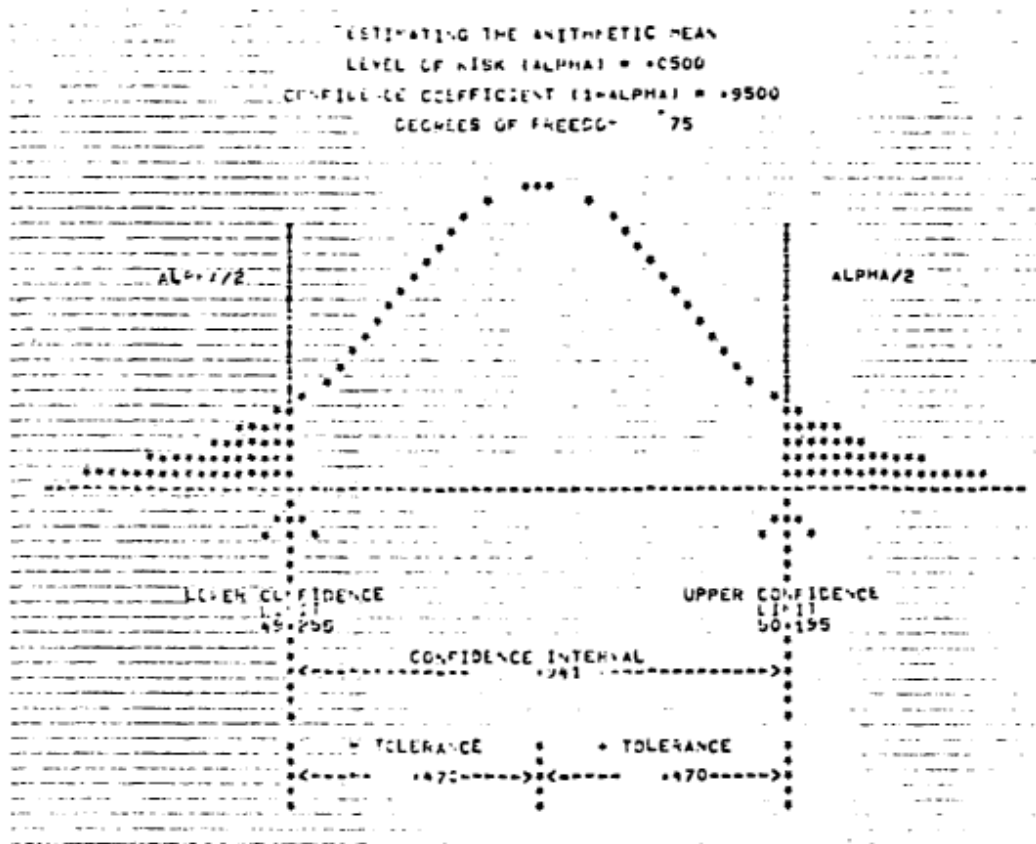CHART 1

ESTIMATING THE ARITHMETIC MEAN

CHART 2

```
                    ESTIMATING THE ARITHMETIC MEAN
                   LEVEL OF RISK (ALPHA) = .0500
                   CONFIDENCE COEFFICIENT (1-ALPHA) = .9500
                   DEGREES OF FREEDOM    75
```

ALPHA/2 .... ALPHA/2

```
            LOWER CONFIDENCE              UPPER CONFIDENCE
            LIMIT                         LIMIT
            49.855                        50.195

            CONFIDENCE INTERVAL
            <-------------- .341 -------------->

              - TOLERANCE        + TOLERANCE
            <----- .470 ----->  <----- .470 ----->
```

When this report is submitted to his professor, the student is supplied with instructions with regard to accessing the next exercise. In exercise 2 the client specifies the Level of Risk, the maximum Tolerance and, unlike exercise 1, the sample size. These information bits are punched according to the instructions and submitted to the computer center. The output, as before, is the detailed listing of the sample together with the ingredients for computing the mean, standard deviation, and standard error. This set is, as before unique, but different from the earlier set.

The student now is faced with a somewhat different set of circumstances. In the earlier exercise the variance was known, or assumed to be known. In this episode he must calculate the standard error of the mean from the sample, and apply it to the Level of Risk provided by the client to determine the Tolerance. This time when he tests the calculated Tolerance against the specified Tolerance he finds that he has not met the conditions required by the problem statement. He is now faced with three possible choices (Chart 1). He may elect to increase the Tolerance, the Level of Risk or the sample size. The instructions that are forthcoming deny the first two to him. He is then left with the single choice of increasing the sample size. The problem becomes one of how much to increase it. The computer instructions guide him along the path of estimating the population variance from the sample standard deviation and thence to the estimation of the appropriate sample size that will satisfy the conditions specified by the client. By following the now accustomed instructions, the student makes his computations, and submits them in the proper fashion to the computer for evaluation. If his solutions are within the limits prescribed by the system, he continues along the path indicated by Chart 1. If not, he receives a rather curt statement from the computer that he has erred and that he must recalculate and resubmit his work. At the successful conclusion of this section, he is again asked to interpret his work and to make a recommendation as to the course of action his client should pursue. However, this time he makes his report in the form of response to a questionnaire provided by the client which is in fact a 10 question multiple choice quiz. The results of the quiz are reported by the computer to the student in the form of a "check" for his score. He "cashes" his check with his instructor where it becomes a part of his semester score.

Concepts Explored

The student has explored ground much beyond that usually presented in an introductory course to inferential statistics. Moreover, he has probed much of it at his own pace and of his own volition, albeit prodded somewhat by both his instructor and much more importantly, his peers. He has reviewed probability concepts, the concept of standard deviation and variance. Of even greater importance, he has "discovered" the relationships among the Level of Risk, Tolerance, Variance, and Sample Size. He has accomplished all this working almost

independently of his fellow students because each of them has been working on a different set of data making collusion very difficult except in the <u>exploration of concepts</u>. The amount of arithmetic has been minimized; only that necessary to develop the essential elements of the analysis has been required. The tedium usually associated with computation of statistics has been removed. The student has been encouraged, therefore, to think beyond the immediacy of adding, subtracting, multiplying and dividing. He has been involved in developing concepts. This has been the challenge.

Further Developments

Several weaknesses in the program as currently constituted have been revealed during its short life. The first is that the employment of the "t" vs the "Z" pivotal value has been ignored. The second is that only infinite populations have been specified. Both of these deficiencies are scheduled for remedial effort at an early date. Most real life estimates must use the "t" rather than the "Z", and it is important that this modification be made. Often their distinctions are treated lightly in the classroom because of the lack of detailed "t" Tables or because of the added arithmetic that their use generates. The computer removes these restrictions. The use of finite populations, especially those that are relatively small, are also afforded only token coverage. The student begins to wonder about the usefulness of the techniques he has been learning when the usual calculation of the minimum sample size turns out to be larger than the population with which he is concerned.

AN EVALUATION

Student feedback is the only evaluative technique so far employed in examining the worth of this system. Its limitations are well recognized, but like so much of that which we label the "real world" it is all we have. A prevalent student attitude toward statistics has been developed earlier, but even a casual observer is aware that it has not surfaced to an important degree in the classes in which this technique has been employed. Much of its absence may easily be attributed to the novel aspects of the program, and it may return when the newness has vanished. Much of its worth will remain, however, because it does offer a unique challenge. It can be kept timely with only slight revisions each semester in the initial output the student receives. The computational tedium has been reduced to a bare minimum, perhaps only that amount remains which a busy manager might wish to check himself after having received a report from his staff. And the use of the computer has also reduced the response time to almost that which the student wishes it to be. The emphasis has been removed from computation to concept. This facet along makes for a worthwhile effort. And this transfer of emphasis puts the student in the realm of doing for the sake of learning, not f or the sake of doing.