Teaching Statistics through Resampling

KEYWORDS:
Computer; Randomization test, Bootstrap ;Simulation.

Chris Ricketts and John Berry
Centre for Teaching Mathematics,
University of Plymouth, UK.

Summary
This paper describes experiences of teaching statistics without mathematical theory but using computer-intensive re-sampling methods. The method is relevant to statistics teaching at all levels.

◆ INTRODUCTION◆

MANY students of statistics, especially those who are not studying mathematics, have difficulty with the mathematical formulation of statistical methods. The theory of statistics is largely based upon the theory of probability distributions and sampling distributions of parameters such as the mean or test statistics such as chi-squared. Because each approach uses a different formula, many students have difficulty in choosing the right formula to use. We have recently investigated the use of a computer-intensive approach to teaching statistics that is not formula-based.

Computers have had a tremendous impact on the teaching of statistics, especially at post-16 level. Statistics packages are used to carry out the relatively tedious calculations required in many estimation and testing procedures. However, this is doing no more than using the computer package as a very fast calculator. The computer-intensive approach, in contrast, uses the computer to mimic the real-life sampling (and repetitive re-sampling) which gives rise to such things as the Central Limit Theorem, confidence limits and hypothesis testing. There are no formulae involved - these are replaced by a sampling algorithm that reflects the underlying statistical and probabilistic process. The only understanding that is required is to be able to interpret a histogram and to be able to generate a sampling process.

Computer-intensive methods are becoming increasingly popular as computers become faster and cheaper. The commonly used methods are randomization testing and bootstrap estimation. There are few packages available and one of the drawbacks of using these methods to teach probability and statistics is that programming is required. However, the recent development of a new package *Resampling Stats* in the USA and its competitive pricing policy now makes this approach feasible for most schools and colleges. This package has been developed by The Resampling Project led by Julian L. Simon, a professor of business administration at the University of Maryland. For a review of the work of the Project see Peterson (1991) and Simon and Bruce (1991).

In this paper we report on the use of this approach to a mixed experience class in Higher Education, many of whom had not previously studied statistics. It is important to note, however, that the approach can be used either as a substitute for traditional theory or as an aid to understanding it, and is relevant to all levels of probability and statistics teaching, from the National Curriculum to Higher Education.

◆HYPOTHESIS TESTING EXAMPLE◆

We will illustrate the re-sampling approach using one of the most difficult concepts in statistics, but also one that is taught at a variety of levels hypothesis testing. We will take as our example the comparison of two means. The data is from problem 8.17 in ‘Understanding Statistics’ by Ott and Mendenhall (1985).

When testing the efficacy of a drug to reduce blood pressure, 12 rats were chosen and the drug administered to 6 rats the treatment group chosen at random. The other 6 rats - the control group - received a placebo (an inert substitute). The drop in blood pressure (mmHg) observed in each group was:

Control group
9, 12, 36, 77.5, -7.5, 32.5 (mean 26.6)
Treatment group
69, 24, 63, 87.5, 77.5, 40 (mean 60.2)

We are interested in the question of whether the observed difference between the groups is 'significant' or whether it is the sort of difference we might observe by chance in any two samples from an untreated group.

The traditional approach to this problem would be to use the Student *t*-test. In this approach the student would calculate the difference between the means $\bar{x}_2 - \bar{x}_1$ and the standard error of that difference

$$\sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

Then the ratio of the difference between the means divided by its standard error is used as the test statistic (the value is 2.16). The student would look up this value in the tables of the Student *t*-distribution with 10 degrees of freedom to obtain the probability of getting a test statistic this extreme under the null hypothesis that treatment has no effect and that the two samples merely represent two samples from the same distribution.

Using a computer package such as MINITAB removes the need for students to use formulae correctly, but still faces them with deciding on which test procedure to use, whether or not to use the pooled estimate of the variance and whether the alternative hypothesis is two-sided or one-sided and, if the latter, in which direction.

The MINITAB commands required to complete this analysis are set out in Fig. 1, together with the output in Fig. 2.

It is worth noting at this stage that, whether the students do the calculations by hand or using a computer, the analysis is based upon the assumption that the data are two random samples from a Normal distribution with unknown standard deviation σ. The mathematical derivation of the test is beyond the understanding of the students and they are required to interpret the test statistic *(T*= 2.16) and the P-value of 0.028.

```
MTB >set cl
DATA> 9 12 36 77.5 -7.5 32.5
DATA> end
MTB > set c2
DATA> 69 24 63 87.5 77.5 40

DATA> end
MTB > name c1 'control' c2 'treat'
MTB >two sample 'treat' 'control';
SUBC> pooled;
SUBC> alternative + 1.
```

Fig.1

```
   TWO SAMPLE T FOR treat VS control
            N     MEAN     STDEV     SE
treat       6     60.2     23.9      9.7
control     6     26.6     29.7      12

95 PCT CI FOR MU treat - MU control: (-68, 1.1)
ITEST MU treat = MU control (VS OT): T = 2.16
P = 0.028 DF = l0
POOLED STDEV = 26.9
```

Fig 2

```
'Randomization test for the difference between two
means 'Problem 8.17 pp. 242-3 from
Ott & Mendenhall

copy (9 12 36 77.5 -7.5 32.5) control   'control group
copy (69 24 63 87.5 77.5 40) treat      'treatment group
concat control treat all    'all data
repeat 99                   'try 99 randomizations & include
                            'the observed value to make 100
shuffle all new             'randomize the data
take new 1,6 cnew           'call the first 6 control
take new 7,12 tnew          'and the other 6 treatment
mean cnew cbar              'mean for new control group
mean tnew tbar              'mean for new treatment group
subtract tbar char diff     'and calculate the difference
score diff Out              'store all the differences

end

mean control char   'calculate the OBSERVED difference
mean treat tbar
subtract tbar char diff
score diff out

histogram out       'look at the randomization distribution

count out>diff big  'and the probability of getting a value
divide big 100 prob 'at least this big
print diff prob
```

Fig. 3.

The computer-intensive approach to the problem stems from Pitman (1937) and gave rise to the well-known non-parametric approach to this problem. The computer-intensive approach has only recently become feasible as the cost of hardware and software has dropped. Most importantly, the computer-intensive approach does not rely upon the distributional assumptions of the Student *t* test, nor does the student have to remember any formulae. The approach is as follows.

Assuming that the treatment has no effect, each of the observed drops in blood pressure is equally likely to have come from either group. We can therefore Fig. 2. generate the sampling distribution of the difference between the means by randomly allocating results to each group a large number of times. This process approximates the distribution we would obtain by completely generating all 924 permutations of the data. A simple sampling scheme is

i) write each result on a card
ii) shuffle the cards
iii) deal out the first 6 cards as the control group
iv) use the remaining 6 cards as the treatment group
v) calculate the means of each group and the difference between them
vi) repeat steps i)-v)
vii) produce a histogram of the results from vi)
viii) the probability of getting a difference as great as that observed is the proportion of values in the histogram that are at least as big as the observed difference.

A program in the *Resampling Stats* language is given in Fig.3.
The general syntax is

command <input vector(s)> <output vector>

This language has the advantage of mimicking the sampling process in steps i) - viii) above. The comments to the right of each command make the program self-explanatory.

The resulting histogram is shown in Fig. 4. The 99 randomizations took less than 2 seconds on an IBM-compatible 286 computer. Only 2 of the values in this histogram (one of them being the observed value itself) are at least as large as the observed difference of 33.6.

The traditional Student *t* test, with all its assumptions and formulae, gives a one-tailed P-value of 0.028. The re-sampling program above gives a P-value of 0.02 for 99 randomizations and a value of 0.022 for 499 randomizations. All these tests imply that the treatment mean is 'significantly' higher than the control group mean, so that the new drug is effective in reducing blood pressure.

The difference between the two approaches is that the computer-intensive approach makes the sampling distribution of $\bar{x}_2 - \bar{x}_1$ explicit and visible. The students can see how unusual the observed value is. The mathematical approach, on the other hand, uses a formula to transform the difference $\bar{x}_2 - \bar{x}_1$ to a standard distribution from which the P-value can be derived.

◆ STUDENT ATTITUDES ◆

This re-sampling approach was taught to a class of 42 students on a Mathematical Modelling module, some of whom were first year undergraduates, some were second-year undergraduates and a few were third-year trainee teachers. In three 2-hour sessions using *Resampling Stats* we covered

- simulation methods for probability distributions such as the Binomial and computer intensive solutions to the Birthday problem and others;
- confidence limits using the bootstrap;
- hypothesis testing using randomization tests.

In retrospect, this was over-ambitious! Nevertheless, we got some very positive responses from the students, especially the less mathematically able. The students were asked to compare the resampling method with the mathematical method of solving the confidence limit problem. Two comments of particular note were:

(a) "The mathematical solution looks frightening at a first glance but after reading it again it became clearer to me as I had not done something about it before. For me, this solution (using resampling) is easier to understand and if I had no previous knowledge of confidence limits, the resampling

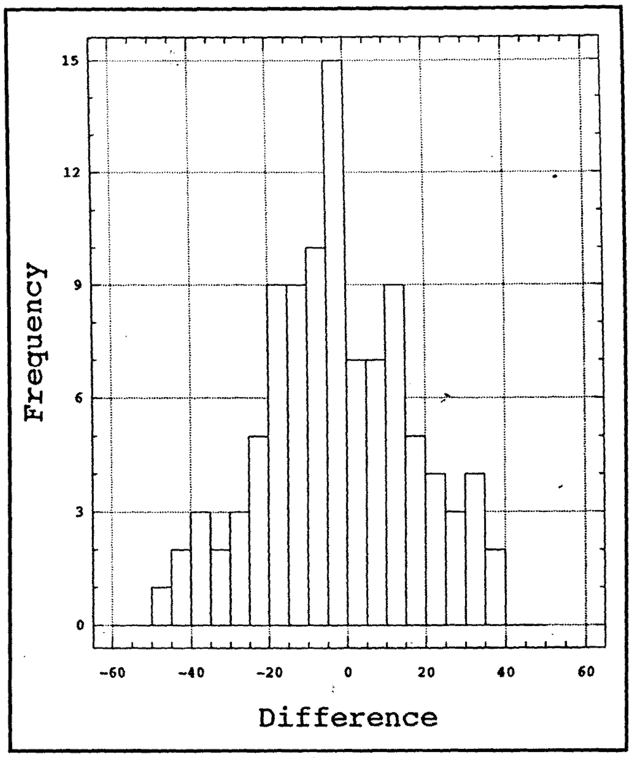

Fig. 4. Histogram of Difference between means

method may have been a good way to introduce it."

(b) "The resampling method is, in my opinion, far easier. to understand than the mathematical solution. However, having done statistics previously, it is difficult to try and distinguish between the hard and easy parts. A very clear difference, I thought, was that the resampling method makes one feel that we are physically doing it, or actually seeing it physically being done, without having to take any theoretical mathematics into consideration."

◆CONCLUSIONS◆

Much of the development of statistics over the past 90 years has been concentrated upon the derivation of test statistics whose sampling distributions can be derived mathematically. The teaching of statistics has followed a similar pattern. However, while the idea of a sampling distribution is relatively easy to comprehend, the mathematical derivation requires skills beyond most students in the 16-19 age range. In contrast, the computer-intensive approach described here requires no mathematical skills and leads students to a better understanding of sampling distributions. Our experience suggests that it is highly acceptable to students with a range of mathematical abilities. The technology is such that we can now use computer-intensive methods either as an alternative to, or preferably in support of, traditional methods of teaching probability and statistics.

References

Ott L and Mendenhall W (1985), *Understanding Statistics*, Duxbury Press, Boston. Peterson 1(1991), *Pick a Sample,* Science News, **140,** 56-58.

Pitman E J G (1937), *Significance tests which may be applied to samples from any population.* Journal of the Royal Statistical Society, **4, 119-130.** Simon L and Bruce P (1991), *Resampling: A Tool for Everyday Statistical Work*, Chance, **4,** 22-32.

The UK distributor of Resampling Stats is QD Consulting, 68 Station Road, Steeple Morden, Royston, 5G8 ONS. (Phone or fax: 0763-852446).