The Statistician (1995)
44, *No.* 1, *pp.* 45–54

Changing the teaching of statistics

By MICHAEL STUART†

Trinity College, Dublin, Republic of Ireland

[Received January 1994. Revised June 1994]

SUMMARY
Statistics teaching has been dominated by mathematical thinking. Examples of this are given. An alternative problem solving approach to statistics teaching is advocated and contrasted with the mathematically oriented approach. The contrast and the advantages of the problem solving approach are illustrated through examples.

Keywords: Mathematical thinking; Problem solving; Statistical education; Statistical thinking

1. Introduction

Statistics teaching up to now has been heavily dominated by a mathematical agenda, as illustrated by the table of contents of a typical text in elementary statistical methods, whether general or intended for a specific application area. The order of presentation of the topics is determined by the mathematical requirements; the necessary mathematics must be in place before any real applications can be discussed. It will be argued later that, not only is this unnecessary; it is counter-productive for most students.

The primary pedagogical effort involved in most such texts is to make the mathematical content as simple and transparent as possible. In most cases, the essential statistical content is barely recognized and many display a disturbing lack of understanding of the statistical concepts on which the mathematics is built. Data are used primarily as illustrations of computational techniques and formal statistical methods.

By contrast, an approach based on solving real problems in applications areas can effectively provide a useful training in statistical analysis, equip students with a capacity for statistical thinking and provide those with the necessary mathematical abilities with a sensible foundation on which to build a more formally mathematical approach to the subject. It is argued here that our approach to the teaching of statistics, particularly elementary statistics, needs to be changed from an approach dominated by mathematical considerations to one which promotes statistical thinking.

It is important to recognize at the outset that statistics as a subject is distinct from, though critically dependent on, mathematics. It is also necessary to recognize that modern statistics is just as critically dependent on computing and computers for its application and development. The subject has always been dependent on substantive application areas for its very existence. Thus, we may view statistics as founded on data in context, while supported by mathematics and computing. The focus in this paper is on teaching users of statistics, whose interests may include theory and methods as well as applications. The paper is not concerned with teaching theoretical statisticians *per se.*

†*Address for correspondence*: Department of Statistics, Trinity College, Dublin 2, Republic of Ireland.
E-mail: mstuart@stats.tcd.ie

 0039–0526/95/44045

The idea that we need to change the teaching of statistics is not new. A recent discussion is provided by Snee (1993) who also provided references to several reports, workshops and conferences in the USA. An additional such reference (brought to the author's attention by a referee) is Gordon and Gordon (1992). Another stream of activity is the series of International Conferences on Teaching Statistics sponsored by the International Statistical Institute. Also, the International Association for Statistical Education was formed recently as a Section of the International Statistical Institute and held its first scientific meeting in 1993. In the UK, the Royal Statistical Society has an Education Committee whose contributions to the Society's newsletter *RSS News* provide up-to-date information on recent and forthcoming developments.

Although the conclusions reached in this paper are not new, the case has usually been made in general terms. The aim here is to provide concrete support for the conclusions by showing, through real examples, the negative influence, on both students and statistical practice, of the dominance by mathematical thinking and the possibility for improvement through an emphasis on statistical thinking. Some relevant recent discussions include Moore (1993), Box (1994) and a series of letters and editorials in *RSS News*: Greenfield (1991); Nelder (1991); Finney (1991); Box (1993); Nelder (1994).

For brevity, many details of the analyses are glossed over in the three examples presented. In doing this, the author takes the risk of being accused of oversights such as those of which he accuses others. The examples are preceded by a discussion of statistical problem solving and followed by further discussion.

2. Statistical problem solving and teaching

A possible paradigm for statistical problem solving would involve the following sequence:

problem formulation,
statistical design,
data collection,
data analysis,
interpretation,
implementation.

In the standard elementary texts, however, most of these are ignored. Instead, the emphasis is almost entirely on formal methods for data analysis, as if data analysis problems came to the statistician ready made, with appropriate mathematical structure. In this section, there is a brief comment on the pedagogical importance of each element in the sequence. Chatfield (1988) gives a detailed treatment of a similar paradigm.

The pedagogical advantage of introducing problem formulation is that it gives a concrete context for statistical issues and sets up substantive questions whose answers require data and statistical analysis. Some discussion of statistical design, based on case-studies, can have the effect of elucidating the nature and sources of variation. By contrast, in the standard approach driven by mathematical thinking, variability is quickly metamorphosed into probability distributions, involving ideas which many students do not really understand.

The data collection process itself is hardly ever mentioned in the standard texts; it is as if the mere desire for data was enough to produce it. Some discussion of issues such as logistics and costs will contribute substantially towards providing an air of reality in a statistics course. It will also act as a useful preamble to discussions of sample size; almost invariably, such discussions in standard texts focus on error probabilities whose determination is left as a mystery.

Most of the attention in the standard texts is focused on the data analysis phase and, within that, on a set of methods of formal statistical inference, each based on appropriate

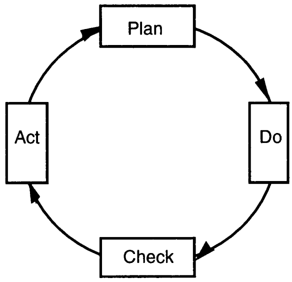

Fig. 1. Shewhart cycle

assumptions about the data. Exploratory analysis and diagnostics are rarely mentioned. Some recent texts include some exposition of 'exploratory data analysis', typically replacing the more sterile 'nonparametric methods' featured in older texts. More often than not, however, this exposition consists of presenting a set of isolated methods, with no attempt to integrate them into an overall approach to data analysis.

Interpretation must be done in the context of the real problem being addressed, with appropriate reservations, and not in terms of the artificial problem entailed in the model assumed at the data analysis phase. Inevitably, initial reservations about interpretations raise questions about the validity of assumptions made in earlier phases. Equally, some involvement with implementing the solution will raise questions about earlier phases. This kind of iteration is a key ingredient of the problem solving approach which must be incorporated in our teaching if we are to turn out productive users of statistics.

The view of problem solving put forward above, although considerably broader than anything suggested in the standard elementary texts, is short term by nature and so rather limiting. For example, the modern approach to quality improvement involves continuous problem solving, following a paradigm such as the Shewhart or plan–do–check–act cycle illustrated in Fig. 1, where the implementation of the solution of one short-term problem stimulates the formulation of a new problem (very often, a variation on the old one). Continuous problem solving may involve using information from any or all of the steps of earlier cycles in any or all of the steps of the current cycle.

3. Some examples

The first example is an abbreviated version of a case-study in the implementation of the problem solving paradigm mentioned in Section 2, followed by criticism of an alternative analysis dominated by mathematical considerations. Next, there is a brief discussion of a text-book example where the authors' own practical advice on model criticism is ignored in favour of spurious mathematical simplicity. Both of these examples feature the use of multiple linear regression. A third example is concerned with an alternative to the standard analysis of two-way frequency data (or 'contingency tables') and concludes with an alternative approach to the introduction of elementary probability.

3.1. *Predicting postal sales*

In 1983, the Irish postal service, formerly operated by a civil service department, was semiprivatized. The new company operating the service, An Post, commissioned consultants to produce sales forecasts. An Post provided historical data on annual postage-stamp sales and meter sales for the years 1949–83, inclusive. For brevity, we concentrate on the former here; a complete analysis would recognize the links between the two. Any sensible approach

Fig. 2. Annual stamp sales, 1949–83 (millions of standard stamp equivalents, i.e. total revenue divided by the nominal price of a stamp for a standard sealed letter for internal delivery)

to this problem will start by looking at the data supplied. Fig. 2 shows a time series graph of stamp sales. The larger jumps correspond to stamp price changes. The dip in 1979 was due to the lengthy industrial dispute in the postal service in that year.

A consideration of what else might influence stamp sales suggested stamp prices, prices of alternative products and general economic growth. Multiple regression was suggested as an appropriate model to link sales to such explanatory variables. Questions relating to the measurement of these variables led to choosing gross national product (GNP), real letter price (RLP) (i.e. the price of a stamp for a standard size local delivery letter, deflated by the consumer price index (CPI) and real phone charge (RPC) (i.e. the price of a local telephone call, also deflated by the CPI) as variables which might explain the observed variation in sales. The

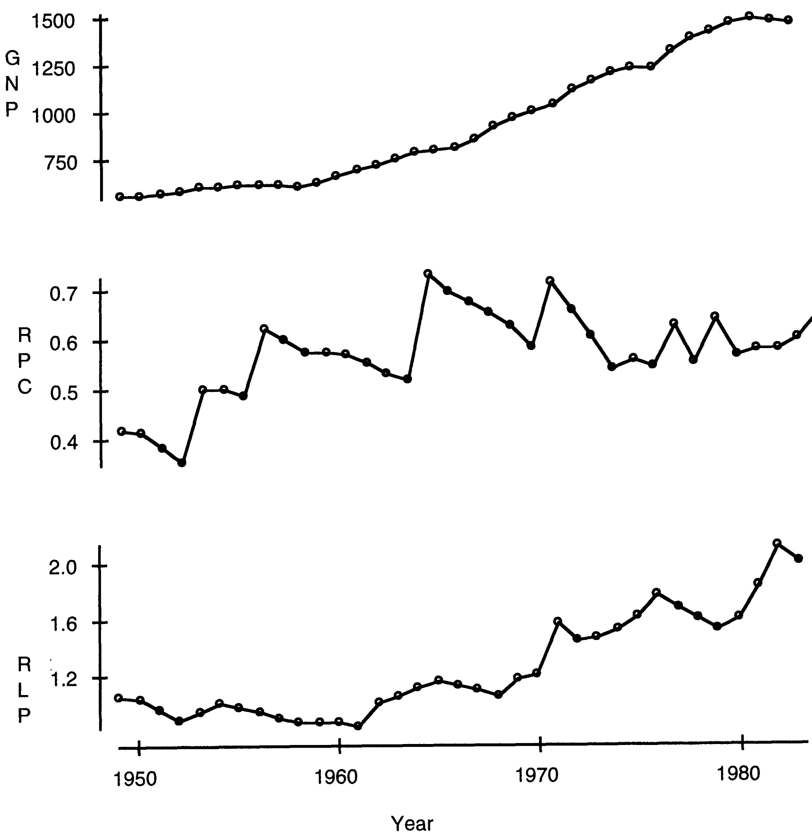

Fig. 3. GNP, RLP and RPC, 1949–83

raw price data were immediately available to An Post; GNP and CPI data were easily acquired from the Central Statistics Office and the price data were adjusted via the CPI. Again, a graphical view of these variables, as in Fig. 3, suggests interesting correspondences. Among them is the fact that sales is a decreasing function of GNP.

Many statistical lessons may be learned from this model formulation and data collection process. Among issues that arise are the definition of GNP and its use to represent general economic conditions, the adequacy of the price of a 'standard stamp' in representing varying prices of a variety of postal services, the definition of the CPI and its use as a 'deflator'. Also worth discussing is the apparently negative relationship of sales to GNP and the role of meter sales in explaining this; such a discussion could signal a future consideration of conditional relationships (Simpson's paradox) and causality. Yet most statistics texts considering such a forecasting problem would ignore most of the lessons. Instead, the problem would arise in the context of a discourse on multiple regression, as an illustration of the application of the technique to ready-made 'dependent' and 'independent' variables.

The traditional approach to teaching multiple linear regression introduces the model and then derives parameter estimators, their properties, relevant tests and analysis of variance, and then introduces an example to illustrate the procedures and calculations. Typically, nowadays, all this activity concludes with computer output which invariably includes an analysis-of-variance table whose immediate relevance is dubious at best. Many recent texts leave it at that. A problem solving statistician, by contrast, would proceed to model validation, for example, by plotting residuals against fitted values, resulting as shown in Fig. 4.

The consultants interpreted the pattern in Fig. 4, having allowed for the known outlier in 1979, as reflecting the error variance being an increasing function of the dependent variable. Accordingly, they chose to use log(sales) as the dependent variable in a second regression. They also used logarithms of all the independent variables. This had the virtue of allowing them to interpret the resulting regression coefficient estimates as elasticities. Having found a statistically significant value for the Durbin–Watson statistic in their first fitted regression,

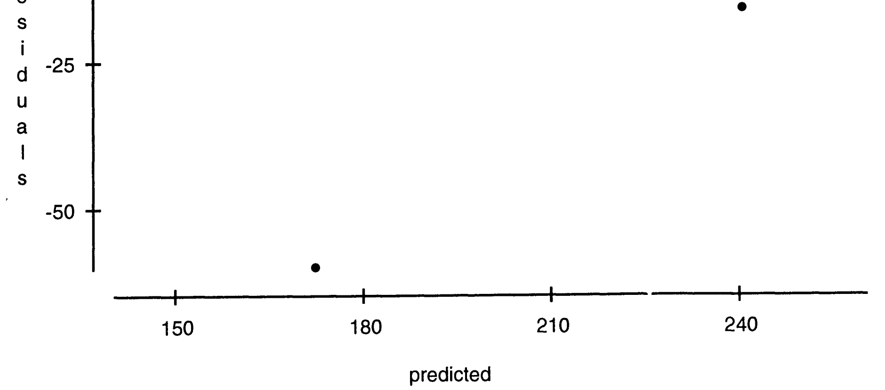

Fig. 4. Residuals *versus* fitted values, full model

they incorporated first-order autoregressive errors in the second fit. They pronounced the resulting fit satisfactory.

An alternative assessment of Fig. 4 is to suggest that, apart from the outlier, there are two groups of residuals, each reflecting a homogeneous error pattern but with different standard deviation values. The points on the left-hand side of the diagram, with fitted values smaller than 200, correspond to the years after 1970; those on the right-hand side to the years 1970 and before. (Recall from Fig. 2 that there was a steady decline in sales over most of the period). Fitting separate models to the early and late data produced, after allowing for previously masked outliers, satisfactory fits in each case with, however, statistically significant 'before-and-after' differences in both linear coefficients and standard deviations. A conversation with the clients suggested that there were economic factors which would explain the change. However, this issue is not pursued here.

It seems clear that the 'two regressions–independent errors' model is at least as plausible as the 'increasing variance–autoregressive errors' model. It may be that the consultants in this case were too much influenced by mathematical thinking. Although the 'constant elasticity' model allows a mathematically convenient interpretation of the coefficients, it is clearly inappropriate here. Also, the apparent significance of the Durbin–Watson statistic is very probably spurious. The Durbin–Watson statistic frequently reacts to things that are quite different from autoregressive errors, e.g. a shift in some or all of the linear parameters such as occurs here.

If we accept the conclusion that two regressions are needed to fit these data, we would naturally use the later data as a basis for projection. The model chosen to fit these data was used with values of the explanatory variables predicted (in 1983) by the Central Bank of Ireland to give predicted stamp sales for 1984 and 1985. The predicted values are shown in Table 1, along with the values predicted by the consultants' chosen equation and the actual sales for those years. Much more of pedagogical value can be derived from this case-study. However, the point here is to contrast the two approaches to analysis as described above.

3.2. *Predicting job completion times*

An example in Bell and Newson (1987), pages 331–333, further shows the danger of being misled by the mathematical simplicity of a model when available statistical diagnostics indicate its inadequacy, exemplifying the unhelpful influence of mathematical thinking in much statistics teaching. The example was concerned with the length of time required to complete orders for various products manufactured by a metal fabrication company. Full background details, including the data, are given by Bell and Newson (1987), so only the essential points are mentioned here.

In the example, the researchers regressed the dependent variable, job completion time, on four independent variables, number of pieces per job, number of operations per piece, their product (the number of operations per job) and an indicator variable which distinguished 'rushed' orders from 'normal' orders. Noting that the t-values for number of pieces per job, number of operations per piece and the indicator variable were statistically insignificant, they argued that the simple linear regression of job completion time on the number

TABLE 1
Predicted and actual stamp sales

	1984	*1985*
Actual sales	164	172
Author's predictions	164	170
Consultants' predictions	167	178

of operations per job provided a satisfactory predictor. This analysis had the virtue of producing a simple and logical model and of seeming to show that the practice of 'rushing' some orders had no real effect.

Unfortunately, the fit was heavily influenced by one outlier, not discovered by the researchers, and, when this was removed, two further outliers were revealed which suggested that large orders had very variable completion times. Also, the simple model was no longer satisfactory. In particular, it was clear that rushing orders was effective in reducing the job completion time. In a problem solving environment, this conclusion could lead to a discussion of the possible effects of rushing jobs on the quality of the product, whether more effective job management could improve job completion times for all jobs, etc.

This statistical diagnosis, revealing the outliers, was the result of inspecting a plot of residuals against fitted values. This practice is advocated and illustrated by Bell and Newson earlier in their book but not used by them in this simple but interesting case-study. Here again it appears that they were too strongly influenced by mathematical thinking, even to the extent of ignoring a statistical thinking practice advocated by themselves.

It is worth noting in this context that all the global statistics such as s, R^2, F, t and Durbin–Watson require exception-free data and will be meaningless and possibly misleading otherwise.

3.3. *Cross-tabulated frequency data*

An area where the text-books seem to be totally dominated by mathematical thinking is the analysis of two-way contingency tables. The almost universal approach is to test the null hypothesis of independence of row variable and column variable by using the χ^2 test of significance and all the pedagogical effort goes in explaining the mechanics of the test, which may be accompanied by 'typical' computer output. As an illustration, consider a market research study in which interviewees were asked whether they had never heard of a product, had heard of it but had not bought it or had bought it at least once. The results, classified by region, are presented in Table 2. The value of the χ^2-statistic for testing independence is 25, with 4 degrees of freedom, highly significant by most standards.

This analysis, which is as far as most elementary texts go, is grossly inadequate, for several reasons. For one, the display of the data in Table 2 is not immediately informative. To achieve a more informative display, the first step is to recognize that the data are concerned with a *response* variable and an *explanatory* variable. A marketing manager would be interested in the *penetration pattern*, i.e. the relative numbers responding at each penetration level, and how that penetration pattern varies *between regions*. Such questions are more readily answered by inspecting Table 3 in which the penetration pattern in a region is represented by the triple of percentages in the corresponding row.

Note that the rows have been rearranged to emphasize the similarity of the penetration patterns in regions A and C, and their difference from that in region B. If that difference were real, the marketing manager should consider what steps to take to improve the relatively

TABLE 2
Frequency of responses classified by penetration level and region

Region	*Never heard*	*Heard, did not buy*	*Bought*	*Totals*
A	36	55	109	200
B	45	56	49	150
C	54	78	168	300
Totals	135	189	326	650

TABLE 3
Responses at each penetration level, per cent in each region

Region	*Penetration level*			(*Totals*)
	Never heard	*Heard but not buy*	*Buy*	
A	18	28	55	(200)
C	18	26	56	(300)
B	30	37	33	(150)
Aggregate	21	29	50	(650)

poor performance in region B. Note also that the percentages are rounded to whole numbers, following a precept advocated by Ehrenberg (1975, 1982) and others. The whole question of data display, both tabular and graphical, is widely ignored in the standard texts, with some recent exceptions. Mathematics has little to say on this question. The key issues are concerned with communication and perception, areas that are more in the realm of pyschology.

Another major problem with the standard approach is the way in which the χ^2-analysis is introduced. The independence model makes sense to students only if they have understood the concept of independence when it was introduced in the chapter on elementary probability. Conversations with elementary statistics teachers suggest that this is the exception rather than the rule. In fact, many teachers use the contingency table chapter as a second opportunity to introduce the idea of probabilistic independence, with the danger of failing to get across either independence or χ^2-analysis.

The language used in the standard approach to contingency table analysis may also be a source of confusion. The word 'contingency' itself is unfamiliar to many students and therefore a likely source of confusion. More critically, however, the use of the term 'expected frequencies' for the frequencies calculated on the assumption of the independence hypothesis may be counter-productive. Those many students who do not understand the independence model will have difficulty in associating the expected frequencies with it. In any case, the so-called expected pattern is usually so different from any of the observed patterns as to call into question the sanity of the statistician who apparently proposes it as a realistic expectation. In most applications, as in this one, a two-way contingency table consists of the frequency distribution of a categorical response variable classified by a stratification factor of interest. In a sensible study design, the stratification factor is chosen because we *expect* differences. To many, therefore, the statistician's use of the term expected frequencies seems an abuse of ordinary language.

In a problem-oriented statistics course, one might start with the marketing manager's need to assess market penetration, with discussions of issues such as the measurement of market penetration, the design of a questionnaire to allow categorization of respondents, the selection of an appropriate sampling strategy, possible interviewing strategies and their associated costs, and the choice of sample sizes. This may be followed by an analysis of the results, providing an operational interpretation in terms of explaining the response and considering strategic marketing consequences. If needed, the χ^2-analysis can be set up by apportioning row totals according to the aggregate pattern to obtain the so-called expected frequencies on the assumption of only chance differences between regions. Then, instead of the negatively sounding conclusion that we reject the null hypothesis, we can have the positive conclusion that the stratification factor explained much of the variation in the response variable and we have the constructive suggestion to the marketing manager that specific explanations for this be sought.

In practice, data like these rarely occur in isolation; there may be data from previous surveys; there may be other related management information, e.g. actual sales in the regions. A key part of seeking explanations of the pattern in the present data set involves placing it in the context of such additional information and then attempting to make sense of it. Such activity is never discussed in the conventional introductory texts. However, formal statistical inference, which dominates such texts, contributes virtually nothing to such attempts.

There is an added bonus from this approach. The conditional relative frequencies which constitute the penetration patterns and the sampling context can serve as a basis for introducing conditional probability, from which the usual framework of joint and marginal probabilities may be developed, with independence emerging as a special case. It is suggested that this way of introducing formal probability concepts for those who need and can appreciate them is considerably more natural and effective than the standard approach.

4. Discussion

A key development in recent years has been the isolation of the notion of statistical thinking. Moore (1986) described statistical thinking in the following terms:

> 'Statistical thinking embraces the idea of a process, the omnipresence of variation in processes, the explanation of variation . . . , and the need for data about processes'.

This definition is particularly suited to industrial applications and may be too narrow for general purposes. However, it has strong pedagogical advantages. In particular, there is a key pedagogical advantage of thinking in terms of processes instead of populations. In introducing the key concepts of statistical inference, particularly the concept of sampling distribution, the notion of repeated sampling is crucial. If we think in terms of population, then we must talk about repeated sampling from a hypothetical population of all possible values of the measurement made under the same conditions, whereas, using the process idea, we have repeated sampling from an in-control process.

The first requires development of a series of abstract, artificial and obscure ideas which first-time students find extremely difficult. The second is much simpler, much more realistic and can be readily exemplified concretely. It may be more appropriate to introduce population ideas, both real and hypothetical, after students have had some experience of statistical inference in the process setting.

In addition, the focus on process emphasizes real problems that people need to solve. There is a sense of activity implicit in the term. If there is a problem with the process, then that implies wasted activity and the sooner the problem is addressed and solved the better. The alternative more traditional focus on populations puts the emphasis on inactivity, sameness, which does not contribute the same sense of urgency to solve a problem and leads more to a desire merely to analyse the problem, just a subset of the solution.

In contrasting the positive role of statistical thinking with the negative influence of mathematical thinking, there is no suggestion that mathematics is unimportant in statistics. Indeed, many researchers from Hotelling (1940) to Box (1993, 1994) have suggested that statisticians cannot know enough mathematics. The point being made here is that mathematics has had an undue influence both on statistics teaching and on the development of statistics as a subject. The big danger with this is that the mathematical abstraction of a statistical problem invariably ignores some practically important aspects of the problem or, at best, directs attention from them. Typically, the mathematical abstraction of statistical concepts takes them out of their practical setting, transforms them and then attempts to reapply the results, or at least to consider the results applicable. The traditional dominance of statistics by mathematics in this way has disfigured our subject and, because it has also dominated our approach to teaching, even at the elementary level, it has disfigured much of statistical practice. Furthermore, it has prevented much more statistical practice, by putting off the mathematically disinclined among the statistically promising.

An interesting parallel remark is that of Davidson (1992) in a paper in which he argued that real world experience, not prolific publishing, should be a requirement for operations research or management science university teachers. In response to another author's suggestion that 'in teaching OR, even to Ph.D.s, we must transcend our own love of mathematical methodology to make room for additional important matters', Davidson suggests 'that lovers of mathematics methodology be banished from the teaching of OR. Respecters, of course; but lovers, never.'

5. Conclusion

The problems of statistical practice are different from those of mathematical statistics. Statistical ideas (and not their mathematical abstractions) should be the focus of our teaching.

Acknowledgements

Eamonn Mullins provided valuable criticism of various drafts of this paper which led to substantial improvements. The referees made several valuable suggestions; given the subjective nature of the topic, I did not agree with all of them.

References

Bell, P. C. and Newson, E. F. P. (1987) *Statistics for Business with MINITAB, Text and Cases.* Redwood City: Scientific Press.
Box, G. (1993) Scientific statistics. *RSS News*, **21**, no. 4, 1–2.
——— (1994) Statistics and quality improvement. *J. R. Statist. Soc.* A, **157**, 209–229.
Chatfield, C. (1988) *Problem Solving: a Statistician's Guide.* London: Chapman and Hall.
Davidson, H. O. (1992) Practice or perish. *OR/MS Today*, **15**, no. 5, 52–53.
Ehrenberg, A. S. C. (1975) *Data Reduction.* New York: Wiley.
——— (1982) *A Primer in Data Reduction.* New York: Wiley.
Finney, D. J. (1991) Letter to the Editor. *News & Notes*, **18**, no. 1, 2
Gordon, F. and Gordon, S. (eds) (1992) Statistics for the twenty-first century, *MAA Notes 26*. Washington: Mathematical Association of America.
Greenfield, T. (1991) Learning in the works. *News & Notes*, **17**, no. 9, 1.
Hotelling, H. (1940) The teaching of statistics. *Ann. Math. Statist.*, **11**, 457–470.
Moore, D. S. (1986) Planning statistical education for undergraduates: strategy and tactics. *Proc. Am. Statist. Ass. Statist. Educ. Sect.*, 39–42.
——— (1993) A generation of statistics education: an interview with Frederick Mosteller. *J. Statist. Educ.*, **1**, 1.
Nelder, J. (1991) Letter to the Editor. *News & Notes*, **18**, no. 1, 2.
——— (1994) Science—a teaching framework. *RSS News*, **21**, no. 6, 1–2.
Snee, R. D. (1993) What's missing in statistical education. *Am. Statistn*, **47**, 149–154.