*Center for Quality and Productivity Improvement*
UNIVERSITY OF WISCONSIN
610 Walnut Street
Madison, Wisconsin 53705

(608) 263–2520
(608) 263–1425 FAX

*Report No. 146*

# Scientific Statistics, Teaching, Learning and the Computer

George Box

*June1996*

*C Q P I*
UNIVERSITY OF WISCONSIN

Report No. 146

***Report Series***
*in Quality and Productivity*

# Scientific Statistics, Teaching, Learning and the Computer

George Box

Center for Quality and Productivity Improvement
University of Wisconsin-Madison
610 Walnut Street, Suite 575
Madison, WI 53705 USA

*ABSTRACT*

*It is argued that the domination of Statistics by Mathematics rather than by Science has greatly reduced the value and the status of the subject. The mathematical "theorem – proof paradigm" has supplanted the "iterative learning paradigm" of scientific method. This misunderstanding has affected university teaching, research, the granting of tenure to faculty and the distributions of grants by funding agencies. Possible ways in which some of these problems might be overcome and the role that computers can play in this reformation are discussed.*

*KEYWORDS:* *Mathematical Statistics, Scientific Statistics, Iterative learning, Orthogonal arrays.*

**CQPI Report No. 146, June1996**

# Scientific Statistics, Teaching, Learning and the Computer

George Box

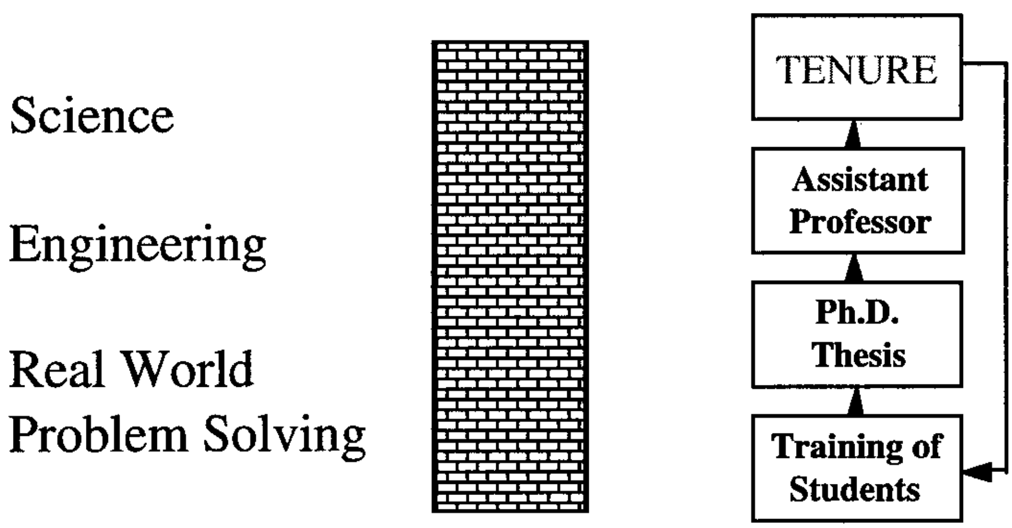
*It is argued that the domination of Statistics by Mathematics rather than by Science has greatly reduced the value and the status of the subject. The mathematical "theorem – proof paradigm" has supplanted the "iterative learning paradigm" of scientific method. This misunderstanding has affected university teaching, research, the granting of tenure to faculty and the distributions of grants by funding agencies. Possible ways in which some of these problems might be overcome and the role that computers can play in this reformation are discussed.*

## 1. MATHEMATICAL STATISTICS OR SCIENTIFIC STATISTICS?

An important issue in the 1930's was whether statistics was to be treated as a branch of Science or of Mathematics. To my mind unfortunately, the latter view has been adopted in the United States and in many other countries. Statistics has for some time been categorized as one of the Mathematical Sciences and this view has dominated university teaching, research, the awarding of advanced degrees, promotion, tenure of faculty, the distribution of grants by funding agencies and the characteristics of statistical journals.

All this has, I believe, greatly limited the value and distorted the development of our subject. A "worst case" scenario of its consequences for teaching, awarding of degrees, and promotion of faculty is illustrated in the flow diagram in Figure 1.

I think it is fair to say that statisticians are not presently held in high regard by technologists and scientists, who feel, with some justification, that whatever it is that many mathematical statisticians have been doing, it has little to do with their problems. Enormous areas of investigation, discovery and development in industry and the universities are thus presently deprived of the refreshing effect of a major catalyst.



It is this view of the irrelevancy of our subject that has perhaps led to the consequences discussed recently by the then-president of the American Statistical Association in an article entitled "Statistics Departments under Siege" (Inman, 1994) in which he speaks of the closings of some university departments and the decline of others. It may still not be too late to reverse this tide and the sea change in computing power that has been transforming our world can help to do this.

## 1.1 INADEQUACY OF THE MATHEMATICAL PARADIGM

The present training of many statisticians renders them competent to *test* a tentative solution to a problem once it has been reached but not to take part in the long, painstaking and often exciting job of discovering that solution. In other words they are not equipped to take part in the process of investigation itself. How can this be?

A purely mathematical education is centered on the one shot paradigm - "Provide me with all the assumptions and conditions associated with some proposition and if it's true I will provide a proof." Not surprisingly this mind-set has also become the paradigm for mathematical statistics - "Provide me with the hypothesis to be tested, the alternative hypothesis and all the other assumptions you wish to make about the model and I will provide an 'optimal' decision procedure." For experimental design this becomes - "Identify the response(s) to be considered, the variables on which they depend, the experimental region to be explored, the model to be fitted, and I will provide you with an alphabetically 'optimal' design."

Perhaps strangest of all is the insistence by some mathematical statisticians that the subject be based on the application of the principle of "coherence". Since this principle can only be maintained within a fixed model (or models) known a priori it excludes the possibility of new departures not mathematically inherent in the premises. Its strict application to the process of investigation would thus bring scientific *discovery* to a halt.

These mathematical straight-jackets are of little use for scientific learning because they require the investigator to provide a priori of all the things he doesn't know.

## 1.2 SCIENTIFIC LEARNING

The paradigm for scientific learning has been known at least since the times of Aristotle (384–322 B.C.) and was further discussed for example by the great philosopher Robert Grosseteste (1175– 1253 A.D.). It is also inherent in the so called Shewhart-Deming cycle: Plan – Do – Check – Act. This iterative inductive-deductive process is not esoteric but is part of our every day experience. For example suppose I park my car every morning in my own particular parking place. As I leave my place of work I might go through a series of inductive-deductive problem solving cycles something like this

| | |
|---|---|
| Model: | Today is like every day |
| Deduction: | My car will be in my parking place |
| Data: | It isn't! |
| Induction: | Someone must have taken it |

| | |
|---|---|
| Model: | My car has been stolen |
| Deduction: | My car will not be in the parking lot |
| Data: | No. It is over there! |
| Induction: | Someone took it and brought it back |

| | |
|---|---|
| Model: | A thief took it and brought it back |
| Deduction: | My car will be broken into |
| Data: | No. It's unharmed and it's locked! |
| Induction: | Someone who had a key took it |

| | |
|---|---|
| Model: | My wife used my car |
| Deduction: | She has probably left a note |
| Data: | Yes. Here it is! |

This iterative process is inherent in all discovery and is, I am sure, part of the long and arduous struggle that creative mathematicians must go through to *arrive* at their propositions – propositions which are eventually published and proved deductively with the elegance and precision of a magician pulling a rabbit out of a hat.

Studies of the human brain over the last few decades have confirmed that, for example, the great mathematician Henri Poincaré and the eminent psychologist William James had long ago suspected: that the brain is divided into two parts constructed to perform jointly this inductive-deductive iteration. For the majority of people the left brain is particularly concerned with *deduction*, analysis and rational thinking, while the right brain is much more concerned with *induction*, pattern recognition and creative insight. A continuous conversation between the two takes place via the interconnecting corpus callosum. Thus the generation of *new* knowledge through investigation must take place as an iterative inductive-deductive learning process. It has the necessary property that different starting points and different investigation routes can lead to success and sometimes to different but equally satisfactory solutions. Obviously this dynamic scientific paradigm cannot be squeezed into any static mathematical formulation; for no one can foretell the route that an investigation will take and the ways in which the assumptions, the responses, the variables of interest, and the experimental regions of interest, will all change as the investigation proceeds. Although the statistician cannot himself supply the necessary subject matter knowledge which the scientist-technologist-engineer will be led to inject at each stage of the investigation, nevertheless he can greatly catalyze the iteration by judicious choice of experiments to explore current possibilities and resolve uncertainties. Also the illumination provided by appropriate analysis and in particular graphic analysis can greatly help the inductive ability of the investigator. While it is an essential part of scientific method to rigorously explore the consequences of assumed knowledge its paramount purpose is of course the discovery of new knowledge. There is no logical reason why the former should impede the latter – but it does.

## 2. STATISTICS FOR DISCOVERY

Past attempts to break out of the limitations imposed by the one shot mind-set have tended to be absorbed and stifled.

Thus some years ago John Tukey and his followers developed and clearly distinguished between exploratory data analysis on the one hand, and confirmatory data analysis on the other. Many statistics departments, at first unsympathetic to these ideas, now assure us that exploratory data analysis is part of their curriculum. But examination often shows that their exploratory data analysis has been reduced to the disinterment and repeated post mortem examination of long dead "data sets" and past experimental designs over which they can no longer have any influence. In these courses it seems unlikely to be mentioned, for instance, that in a live investigation the finding of a suspected bad value should lead the statistician to walk over to the plant or lab where the investigation is going forward to find out what happened. Perhaps the bad value turns out to be just a clerical error, but sometimes what is found can be highly informative. It may lead to certain runs being redone, to new variables associated with abnormal conditions being introduced, and in some cases to new discovery. Whatever happens, is better than a continuance of stationary agonizing.

A second example concerns so-called "response surface methodology". In the late 1940's earlier attempts to introduce statistical design at the Imperial Chemical Industries in England using large preplanned all-encompassing experimental designs had failed. The "one-shot" approach, with the experimental design planned at the start of the investigation when least was known about the system, was clearly inappropriate. The industrial environment where results from an experiment were often available within days, hours, or sometimes even minutes, called for methods matching those of the skilled investigator which allowed for modification of ideas as experimentation progressed. Response surface methods (Box and Wilson, 1951) were developed in consort with industrial experimenters as one means of filling this need. Fractional factorial experiments were used and where necessary augmented to help in the screening and selection of important factors, steepest ascent was used to allow appropriate relocation of the experimental region, sequential assembly of designs was introduced so that designs could be built up to match the simplicity or complexity of the activity in the region under current exploration and so forth. By these means experimentation and analysis were given movement and provided with some of the adaptive properties necessary to learn about the many different aspects of the investigation.

It was this *dynamic* property that was different from what had gone before. My colleagues and I thought of our efforts to develop such techniques as only a beginning and hoped that our work might inspire others to further develop such methods for experimental learning. However we were doomed to disappointment. It is certainly true that sessions on "Response Surface Methods" are a frequent feature of statistical meetings. But I find these sessions most disheartening. Mathematics has once more succeeded in killing the dynamic features of this kind of experimentation and analysis. In particular one listens to many discussions of the generation of fixed designs in fixed regions in known variables with dubiously "optimal" properties.

So one wonders whether and how this state of affairs can be changed.

## 3. CAN STATISTICS DEPARTMENTS BE REFORMED?

If we look at the history of our subject we find, I think, that the genuinely new ideas in statistics have usually come from statisticians who were also scientists, or from teamwork with such investigators. Examples are Gauss, Laplace, Gosset, Fisher, Deming, Youden, Tukey, Wilcoxon, Cox, Daniel, Rubin, Friedman, and Effron. A reasonable inference is that, while it is true that the more mathematics we know the better, we must have scientific leadership. Some of the ways that teaching departments might change are:

a) Previous work in an experimental science should be a pre-condition for acceptance as a statistics student.

b) When this condition is not met, suitable remedial courses in experimental science should be required, just as remedial courses in appropriate mathematics might be needed for students from, say, biology.

c) Evidence of effective cooperative work with investigators resulting in new statistical ideas should be a requirement for faculty recruitment and promotion.

d) Ph.D. theses of mainly mathematical interest should be judged by the mathematics department.

e) If statistics departments find that it is not possible for them to teach scientific statistics, then they should encourage engineering, physical, and biological departments and business schools to do so, instead.

## 4. THE ROLE OF COMPUTATION

The revolution in computer power will, I believe, further catalyze scientific statistics, both in its deductive and inductive phases and will perhaps help to solve our problems. It can also greatly help students to learn about what they need to know.

### 4.1 DEDUCTIVE ASPECTS

There are many ways in which intensive computation can help the deductive phases of learning. In particular it can allow the investigator to look at the data from many different viewpoints and associated tentative assumptions. One recent application where intensive computation is essential is in the analysis of screening designs. It has recently been discovered that certain two-level orthogonal arrays have remarkable projective properties. For example it turns out (Bisgaard, 1987; Lin and Draper, 1992; Box and Bisgaard, 1993) that the twelve run orthogonal array of Plackett and Burman can be used to form a "saturated" design to screen up to 11 factors with the knowledge that every one of the 165 choices of 3 columns out of 11 produces a full $2^3$ design plus a half replicate which is itself a main-effect plan. Thus if, as is often the case in practice, activity is associated with only three or fewer factors, we have the assurance that all main effects and interactions for these factors can be estimated free of aliases. The 12 run design is therefore said to be of *projectivity* P = 3. It can be formerly proved (Box and Tyssedal, 1994, 1996) that many other orthogonal arrays but not all of them can be used to produce designs of projectivity 3. However the number of possible choices of 3 factors quickly increases with larger arrays. For instance while the 20 x 20 orthogonal array of Plackett and Burman can be used to screen up to 19 factors at projectivity 3, there are 969 possible 3 dimensional projections producing at least one $2^3$ factorial design in the chosen factors (in fact 816 of the projections produce a $\left\{\left(2\times 2^3\right)+2_{\text{III}}^{3-1}\right\}$ designs and the remaining 153 projections are $\left\{2^3+\left(3\times 2_{\text{III}}^{3-1}\right)\right\}$ designs). With so many possibilities it is not to be expected that the important factors can usually be tied down in a single iteration. Furthermore the possibility must be taken account of that more than 3 factors may need to be considered. In recent work

(Box and Meyer, 1993) it has been shown how a Bayesian approach may be adopted to compute the posterior probabilities of various numbers of factors being active. The likely combinations are now reconsidered and (Meyer et al, 1994) have shown how intensive computation can select a further subset of experiments to run which maximize the expected change in entropy. After the additional experiments have been run the posterior probabilities can be recalculated and the process repeated if necessary.

### 4.2 INDUCTIVE ASPECTS

The creative ability of the human brain is most stimulated by graphic representation of results. It is by devising creative and interactive graphics to display statistical features that the creative mind of the investigator can be stimulated. If the statistician has little experience of iterative learning he may be involved in a dialogue like the following:

*Investigator*: "You know, looking at the effects on $y_1$ of variables $x_2$ and $x_3$ together with how those variables seem to affect $y_2$ and $y_3$ suggests to me that what is going on *physically* is thus and so. I think, therefore, that in the next design we had better introduce the new factors $x_4$ and $x_5$ and drop factor $x_1$."

*Statistician*: "But at the beginning of this investigation I asked you to list *all* the important variables and you didn't mention $x_4$ and $x_5$."

*Investigator*: "Oh yes, but I had not seen these results then."

### 4.3 LEARNING

Figure 2(a) represents the past pattern for teaching. The students mind is being used here as a storage and retrieval system. This is a task for which it is not particularly well adapted.

**Figure 2(a).** Traditional method of teaching

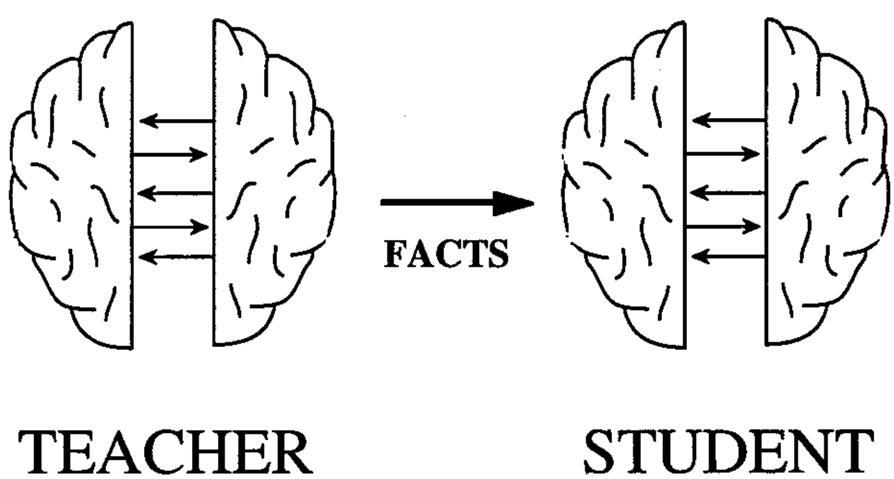
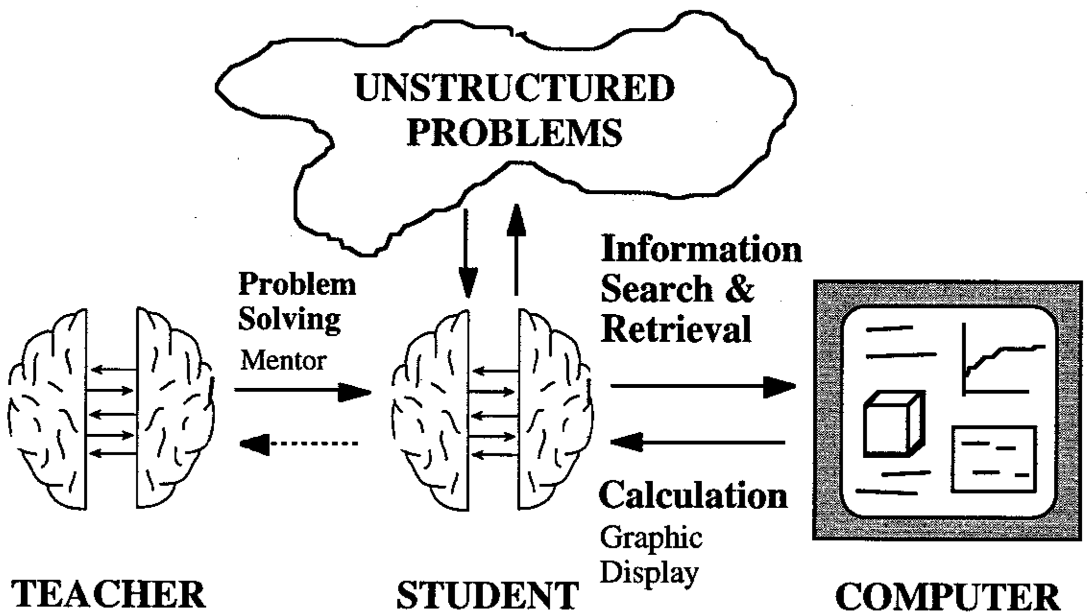



Figure 2(b) shows what I believe will be the teaching of the future. Here the teacher acts as a mentor in training the student in unstructured problem solving, somewhat after the pattern adopted in the medical profession in the training of residents. The computer here plays the part of storing and retrieving information while the human mind is set free to do what it does best and what the computer cannot do, which is to be inductively creative.

**Figure 2 (b).** A model for modern teaching




## ACKNOWLEDGMENTS

This work was sponsored by a grant from the National Science Foundation #DMI-9414765 and to the Universidad de Cantabria, Santander, Spain.

## REFERENCES

Inman. R. L. (1994). "Statistics Departments Under Siege," *Amstat News*, 212 (6).

Box, G. E. P. and K. B. Wilson (1951). "On the Experimental Attainment of Optimum Conditions," *Journal of the Royal Statistical Society*, Series B, Vol. 13. pp. 1-38, discussion pp. 39-45.

Box, G. E. P., Søren Bisgaard and Conrad Fung (1987). *Designing Industrial Experiments: The Engineer's Key to Quality*. Center for Quality and Productivity Improvement; University of Wisconsin; Madison, WI.

Lin, K. J. and N. R. Draper (1992). "Projection Properties of Plackett and Burman Designs," *Technometrics*, Vol. 4. pp. 423-428.

Box, G. E. P. and Søren Bisgaard (1993). "What Can You Find Out From 12 Experimental Runs?," *Quality Engineering*, Vol. 5, No. 4. pp. 663-668.

Box, G. E. P. and John Tyssedal (1994). "Projective Properties of Certain Orthogonal Arrays," *Report #116*, Center for Quality and Productivity Improvement. University of Wisconsin-Madison; Madison, WI. (Submitted to *Biometrika*).

Box, G. E. P. and John Tyssedal (1996). "The Sixteen Run Two-Level Orthogonal Arrays," *Report #135*, Center for Quality and Productivity Improvement. University of Wisconsin-Madison; Madison, WI.

Box, G. E. P. and R. Daniel Meyer (1993). "Finding the Active Factors in Fractionated Screening Experiments," *Journal of Quality Technology*, Vol. 25, No. 2. pp. 94-105.

Meyer, R. Daniel, David M. Steinberg and G. E. P. Box (1994). "Follow-up Designs to Resolve Confounding in Fractional Factorials," *Report #122*, Center for Quality and Productivity Improvement. University of Wisconsin-Madison; Madison, WI. (also to appear in *Technometrics*, 1966.