TEACHING PROBABILITY AND STATISTICS AT UNIVERSITY LEVEL

Invited papers

George Box
Lennart Råde

Contributed papers

Carol Joyce Blumberg
Megan Clark
John D. McKenzie, Jr.
Hristo Nikolov *et al.*
Jenö Reiczigel
Karl Stielau
Roel van Strik
Kamen Velev *et al.*

From **Brunelli, Lina & Cicchitelli, Giuseppe (editors). Proceedings of the First Scientific Meeting (of the IASE).** Università di Perugia (Italy), 1994. Pages 73-87. This pdf file is from the IASE website at http://www.stat.auckland.nz/~iase/publications/proc1993. Copies of the complete Proceedings are available for 10 Euros from the ISI (International Statistical Institute). See http://isi.cbs.nl/sale-iase.htm for details.

WHAT ENGINEERS NEED TO LEARN ABOUT STATISTICS

George Box
Center for Quality and Productivity Improvement
University of Wisconsin-Madison
610 Walnut Street, Madison, Wisconsin 53705, USA

1. Introduction

I will begin by asking three questions and supplying three brief answers which are elaborated on in the rest of my talk. The questions are:

a) Why
b) What
c) How

} should we teach engineers, scientists (and statisticians) about statistics?

I believe the answers are:

a) To catalyze and robustify
b) Methods which catalyze and robustify
c) Engage the student in

} the process of problem solving and scientific discovery

2. Why?

Ensuring that graduating engineers and scientists are familiar with statistical methods of design and analysis is, in the United States at least, an uphill battle (see, for example, Bisgaard, 1991). Remarkably little headway has been made in *requiring* that these techniques are taught to students in engineering and the sciences. This is partly because many scientists and engineers regard the statistics that they *have been* taught as totally irrelevant to the problems they face. I believe that they are often quite right about this.

2.1. *Discovery is an iterative process*

Scientific investigation has two aspects: *discovery* (problem solving) and *testing* the solution. Consider a scientific investigation intended to provide a drug which can cure a particular disease. Such an investigation involves: a) the *discovery* of an effective and manufacturable chemical

substance; b) the *testing* of this substance to ensure its effectiveness and safety for human use.

The process of discovery must be undertaken in the same spirit as a detective solves a mystery and finds a criminal. It is necessarily a *sequential iterative* procedure.

Testing the final product is a much more formal process. It parallels the trial of the accused within very strict rules of admissible evidence. It is usually a *one shot* affair.

Unfortunately, the modern statistician, who is often also the teacher of engineers and scientists, has frequently been trained only for the role of designer and analyzer of the one shot trials appropriate to testing the solution after the work of discovery has been done. Consequently, at least in recent times, statistics has often not been allowed to play its critical role as a catalyst to the process of discovery itself.

Emphasis (Tukey, 1977) on the importance of exploratory data *analysis* addresses this problem. However, exploratory *inquiry* involves not only data analysis but the whole process of investigation and, in particular, the *sequential use of designs.*

2.2. *Continuous never ending improvement*

We can understand the critical importance of sequential investigation if we consider a central principle of modern quality technology - that of "Continuous Never Ending Improvement". This idea seems at first to be in conflict with the law of diminishing returns. Suppose, for example, you have a response curve like Fig. 1(a) or more generally a response surface defined by $y = f(\mathbf{x})$ and you want to find the levels of $\mathbf{x}$ which maximize y. When you have (nearly) achieved this maximization, shouldn't experimentation stop?

This reasoning applies to a fixed model. But (Box, 1993) in real investigations neither the functional form of the model, nor the identity of the variables $\mathbf{x}$, nor even the nature of the response y is fixed. They *evolve* as new knowledge comes to light. As is illustrated in Fig. 1(b), experimenters must be allowed to learn as they go. Whereas a fixed model leads inevitably to the barrier posed by the law of diminishing returns, the developing model provides for expanding returns and the possibility of *never ending improvement.*

In practice, some level of improvement will be adequate for present purposes and the improvement process will temporarily halt; but this quiescent period will last only until external circumstances again renew the incentive for change.

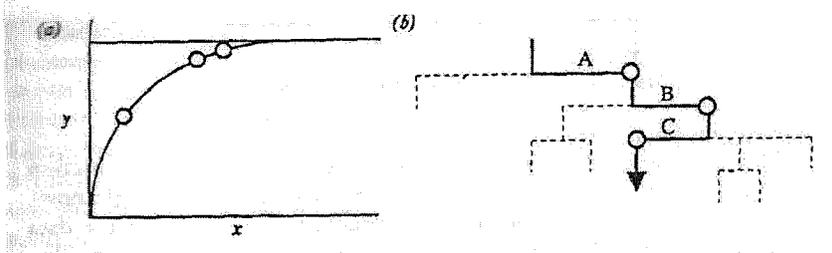


Figure 1. (*a*) *Diminishing returns from the fixed model* $y = f(x)$, (*b*) *Potentially expanding returns from the evolving models* $y_A = f_A(x_A) \rightarrow y_B = f_B(x_B) \rightarrow y_C = f_C(x_C)$. *Dotted lines are "roads not taken".*

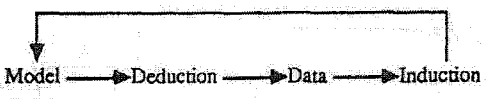
3. What?

The data needed to solve a problem may, at different stages of investigation: a) *already exist* - for example in the library, or in past operating records; b) be obtainable by *observing* but not interfering with the operation of the process; c) need to be *generated* by experimental intervention.

3.1. *Data from passive observation*

We can regard any operating system (whether it be a system for admission to a hospital or a system for producing transistors) as continually generating potentially useful information - rather as a wireless transmitter transmits radio signals. But just as radio receivers are needed to hear radio signals so tools of analysis are needed to understand what the process has to tell us about how it can be improved. The simpler (graphical) techniques such as those employed in Ishikawa's (1976) seven tools are of particular importance. This is because not only can the engineer use them, but s/he can also teach the whole workforce how to use them. The idea is illustrated in Fig. 2 which shows the "seven tools" plus a few more.

Whether by such simple methods or by more sophisticated techniques, effective problem solving necessarily requires a sequential approach. In such a strategy each step uses information gained at previous steps to follow an iterative course.



This process of learning as you go involves a constant interaction between statistical considerations and subject matter (engineering) know-how. As the investigation proceeds it frequently turns out that the data suggests ideas (new variables, different responses, new levels of variables, etc.) that were *not* in mind the beginning of the investigation.

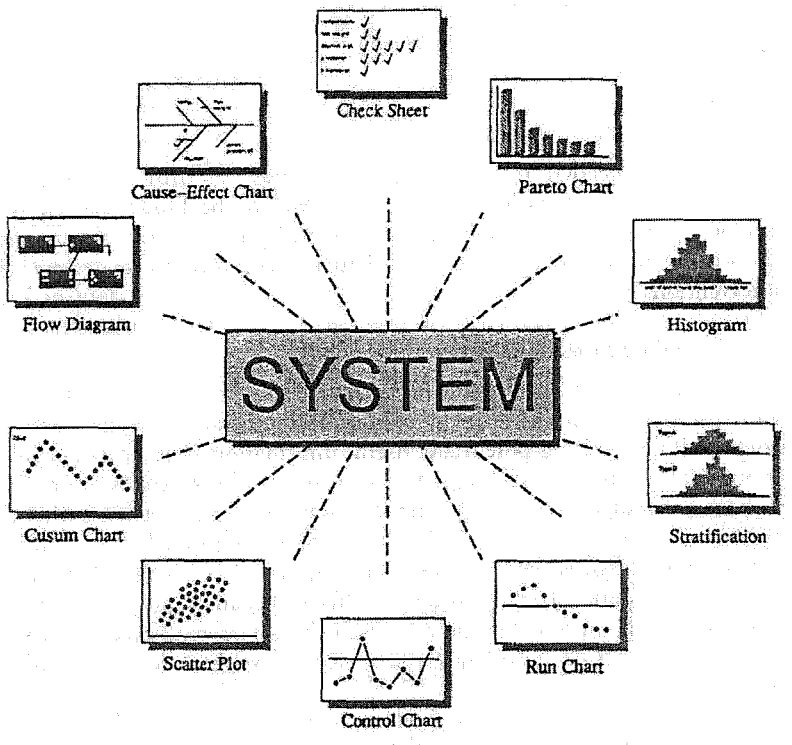


Figure 2. *Some simple statistical tools for quality improvement by the workforce. From top left: Flow Diagram, Cause-Effect Chart, Check Sheet, Pareto Chart, Histogram, Stratification, Run Chart, Control Chart, Scatter Plot, Cusum Chart.*

Statisticians who have been trained only to run one shot trials may be uncomfortable with the idea that they should teach engineers that scientific investigation requires an indeterminate and flexible model, the responsibility for whose evolution must be shared with the experimenter. Certainly in the United States, the role of scientific iteration in statistics is usually dealt with by pretending it doesn't exist and discussing only one-shot investigations. By doing this statistics can be rigorously mathematicized. Unfortunately the process of scientific induction which cannot be modeled mathematically is the only way in which truly *new* ideas can be introduced. By cutting a living process of investigation in two, you kill it.

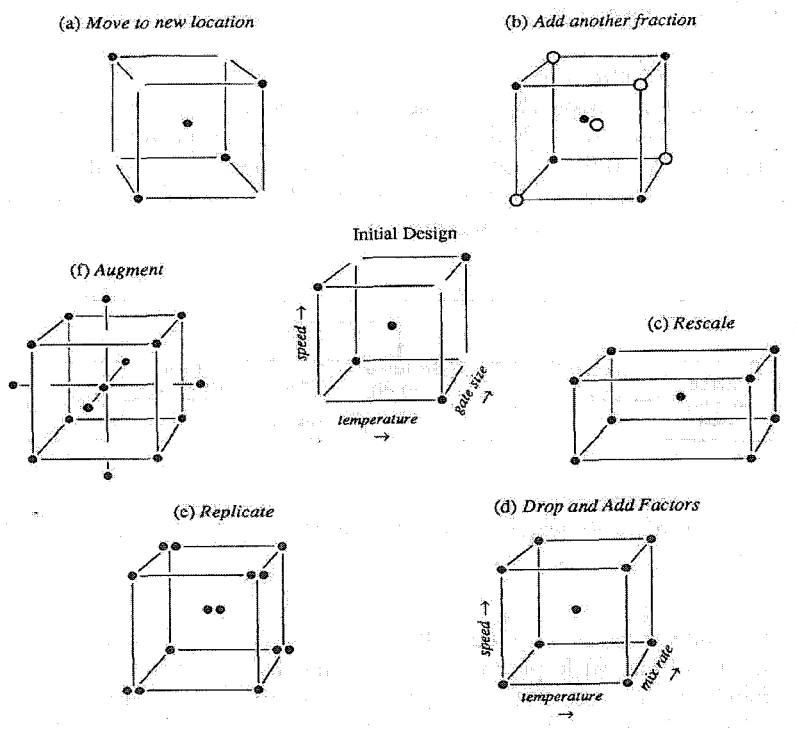


Figure 3. *Illustration of sequential experimentation for the three variable case. Depending on the results from the initial design and subsequent designs, various different courses might be taken.*

3.2. *The use of sequential experimentation*

For problems that cannot be solved by using data already available or obtainable from passive observation of the process designed experimentation is needed. At the beginning of an investigation the experimenter knows least about the *identity* of the important variables, the *location* of the experimental region of interest, the appropriate *scaling* and *transformation* of the variables, the degree of sophistication required to *model* the system and so forth. Thus "one-shot" experimentation which attempts to cover all bases with a single large experiment planned at the beginning of the investigation is likely to be extremely inefficient. Sequential experimentation with associated sequential assembly of designs which at each stage builds on information already obtained is usually much better.

3.3. *Informed extrapolation*

It must be made clear to the student engineer that the scientific process, involves not only the investigational iteration shown on the left of Fig. 4 but also *informed extrapolation* indicated on the right of that figure. Such extrapolation can be from the small scale to the full scale, from one location to another, from one investigator to another, and so on. As has

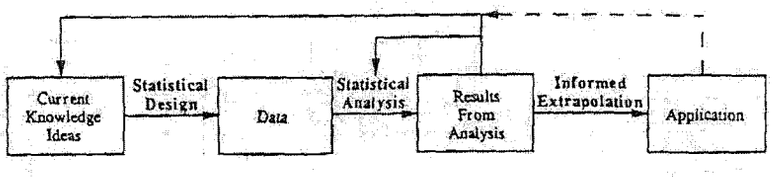


Figure 4. *Feedback and linkages in the process of problem solving and in application of the results.*

been emphasized by Deming (1950, 1986), except in enumerative studies the final link with practice is not made using statistics or formal probability. It is made by using technical judgment. It is important to understand however that the *basis* for this extrapolative technical judgment can be very strong or very weak depending on how the investigation was conducted. Consequently, although no absolute guarantees are possible, by taking certain precautions in the design process we can make the job of informed extrapolation less perilous. The precautions for strengthening the analytical and judgmental links that connect investigation and application

involve issues of *philosophy, analysis,* and *design* and have been discussed more fully elsewhere (Box, 1993). Briefly we need to consider:

Philosophy	a) The investigational process as an iteration
	b) The state of control and null behavior of the system
Analysis	c) The contribution of "distribution free" analysis
	d) The need for exploratory data analysis
	e) The purposes of planned experimentation
	f) The role of randomization and blocking
Design	g) The advantages of comparative experimentation
	h) The rationale of factorial and orthogonal design
	i) The purposes of robust design

(*a*) *Nature of the investigational process.* Iterative investigation, of the kind illustrated in Fig. 1(b) and Fig. 4, itself makes an important contribution to extrapolability.

(*b*) *State of control and null behavior of a system.*The IID state is one that does not occur naturally but can sometimes be approximately induced artificially, for example, by a steady elimination of the larger disturbing factors and/or by feedback control (see, for example, Box and Kramer, 1992). Whether any real process has ever existed in the IID state seems dubious. While every effort must made to ensure that an experimental system is in the best state of control we can get, we would be sanguine indeed to rely absolutely on an assumed state of control to make our experimental conclusions valid.

(*c*) *"Distribution free" analysis.* The role of deductive mathematical analysis in providing a wider base for extrapolation has been greatly overrated. In particular, the so called "distribution free" tests, although free of the assumption of normality, continue to make the same devastatingly dangerous (see, for example, Box and Newbold, 1971) *distributional* assumption that errors are independent or at least interchangeable.

(*d*) *Exploratory data analysis.* Such analysis is of great value at every phase of an investigation, both for providing clues to unexpected phenomena and for warning of situations where standard analysis will be misleading.

(*e*) *Need for planned experimentation.* Fisher, encouraged by Gosset, realized in the 1920's that it would be necessary to *design* experiments to address specific questions of interest. Two concepts were of particular importance - minimizing experimental error and maximizing extrap-

olability. Fisher's (1935) principles of planned experimentation are designed to solve these problems simultaneously.

(*f*) *Randomization and blocking.* Randomization makes the statistical conclusions robust against any kind of non IID disturbance and in particular against a non-stationary disturbance (see, for example, Box, 1990).

(*g*) *Comparative experimentation.* Experiments which most strain extrapolation are those concerned with an *absolute* measurement. Comparative experiments, from which you wish to learn if A and B are different and by how much, are less troublesome. Whenever possible, therefore, a problem of absolute measurement should be transformed into one of relative measurement. For example, in the manufacture of certain automobile parts, a robot grasps each item as it comes off the line and makes a series of measurements. However, these measurements are not compared with standard values. Instead, after measuring each manufactured item, the robot moves back to measure a standard "perfect part" which is available for continuous reference. The *differences* in the measurements of the manufactured part and the standard part are used to decide whether or not the part is in conformance. The advantage of this procedure is that you do not have to have the robot in a perfect state of calibration - any bias will be equally reflected in both the measured part and the reference part.

(*h*) *Factorial and orthogonal experimentation.* Factorial experiments are simultaneously statistically efficient and provide estimates of interactions; in addition they can be run so that the advantages of randomization, blocking and comparative experimentation are maintained. But Fisher (1935) also had in mind the questions of extrapolation and robustness. He remarked "(extraneous factors) may be incorporated in (factorial) experiments designed primarily to test other points; with the real advantages that, if either general effects or interactions are detected, that will be so much knowledge gained at no expense to the other objects of the experiment; and that, in any case, there will be no reason for rejecting the experimental results on the ground that the test was made in conditions differing in one or other of these respects from those in which it is proposed to apply the results".

These ideas were extended further by Youden (1961a,b) who was then working at the Bureau of Standards to develop what he called "rugged" methods of chemical analysis. His experiments used *fractional* factorial designs (introduced by Finney, 1945) to further increase the number of extraneous factors that could be tested.

(*i*) *Robust design.* Thus the concept of using statistics to design a product that will operate well in the conditions of the real world clearly

has a long history going back at least to Gosset in the early part of the century. Early *industrial* examples are due to Morrison (1957) and to Michaels (1964). We owe to Taguchi (1986) demonstration of the wide industrial importance of these robust design ideas.

In the courses we teach engineering students at Madison, we discuss all the above points which impinge on extrapolation but, in teaching the section on robust design, we do not employ Taguchi's techniques (Box, Bisgaard and Fung, 1988). Instead we use what we believe are simpler and more efficient methods which, in particular take account of the important points made by Morrison and Michaels. Also, in short courses, we have successfully taught these ideas to engineers and scientists in many parts of the world.

4. How?

Of prime importance to the engineer investigator is the philosophy of the sequential generation of appropriate data. In particular different design approaches are required at different stages of investigation:

Screening designs. When we do not know which of the number of possible factors are the important ones fractional factorial designs and other *orthogonal arrays* (Plackett and Burman, 1946) are of great importance for screening out what Juran has called the *vital few* factors from the *trivial* many.

Empirical modeling. Sometimes, possibly as a result of previous screening or because of previous knowledge, the important variables are believed to be already known but we wish to find out how they affect a particular response or a number of responses. At this stage of experimentation, factorial designs, mild fractions and *response surface methods* (see, for example, Box and Wilson, 1951 and subsequent publications) are particularly important. The models used at this stage are more or less empirical and are often based on polynomials sometimes in suitably transformed variables.

Mathematical modeling. When sufficient physical knowledge of the system is available *mechanistic model building* techniques may be used. Possibly as a consequence of previous empirical experiment, the process functions derived from a supposed physical mechanism are employed. Such relationships are frequently represented by differential equations or integral equations. They often employ numerical methods of solution and non-linear least squares and non-linear design, together with model checking and model discrimination techniques (see, for example, Box and Draper, 1987; Bates and Watts, 1988).

Experimental design and analysis can also have different *objectives*. Among these are:

i) to raise (change) the mean value of some quality characteristic
ii) to reduce the variance
iii) to find conditions which in some sense produce a *robust* product or process.

4.1. *Hands-on experience*

In teaching the art of catalyzing scientific investigation with statistical methods the most vital component is *hands-on* experience. One way of achieving this (Hunter, 1977) is to require individual projects in which students carry through investigations using statistical methods at home and in the lab. In addition some simple experimental device can be employed for demonstration in the classroom. We have found the paper helicopter shown in Fig. 5(a) to be a very convenient means of teaching experimental design and analysis at many different stages of instruction (see, for example, Box, 1992).

In this first example I will show how a "play acting" scenario may be used to teach the class some fundamental ideas. Three student volunteers are needed to play the parts. I will call these Tom, Dick and Mary.

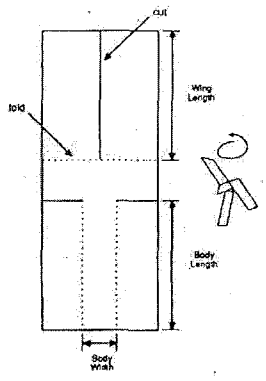
Tom stands on a ladder and drops the helicopter from a height of twelve feet or so while Dick times its fall with a stopwatch. We explain to the class that we would like to find an improved helicopter design that has a longer flight time.

Mean and variation. We start by Tom dropping a helicopter made from blue paper. He drops it four times. We put the results up on the overhead projector and we see that the flight times vary somewhat. This leads to a discussion of variation and to the introduction of the average as a measure of central tendency and the range and standard deviation as measures of spread.

Comparing mean flight times. At this point Dick says "I don't think much of the blue helicopter design. I made this red helicopter yesterday of a different design. I dropped it four times and got an average flight time considerably longer than you just got". So we put up the two sets of data - the four runs made with the blue helicopter and the four runs made with the red helicopter and show the two averages and standard deviations. Then we demonstrate a simple test that shows that there is indeed a statistically significant difference in means, in favor of the runs made with the red helicopter.

Validity of the experiment. Then Mary says "So the difference is statistically significant. So what? It doesn't necessarily mean it's because of the different helicopter design. The runs with the red helicopter were

a)



b)

FACTORS		−	+	EFFECT
Paper Type	(P)	Regular	Bond	0.13
Wing Length	(W)	3.00"	4.75"	0.77
Body Length	(L)	3.00"	4.75"	−0.40
Body Width	(B)	1.25"	2.00"	−0.02
Paper Clip	(C)	No	Yes	0.05
Fold	(F)	No	Yes	−0.10
Taped Body	(T)	No	Yes	−0.15
Taped Wing	(M)	No	Yes	0.17

Random Order	Standard Order	P	W	L	B	C	F	T	M	Flight Time
7	1	−	−	−	−	−	−	−	−	2.5
13	2	+	−	−	−	+	−	+	+	2.9
4	3	−	+	−	−	+	+	−	+	3.5
9	4	+	+	−	−	−	+	+	−	2.7
1	5	−	−	+	−	+	+	+	−	2.0
12	6	+	−	+	−	−	+	−	+	2.3
15	7	−	+	+	−	−	−	+	+	2.9
3	8	+	+	+	−	+	−	−	−	3.0
6	9	−	−	−	+	−	+	+	+	2.4
16	10	+	−	−	+	+	+	−	−	2.6
14	11	−	+	−	+	+	−	+	−	3.2
5	12	+	+	−	+	−	−	−	+	3.7
11	13	−	−	+	+	+	−	−	+	1.9
10	14	+	−	+	+	−	−	+	−	2.2
2	15	−	+	+	+	−	+	−	−	3.0
8	16	+	+	+	+	+	+	+	+	3.0

c)

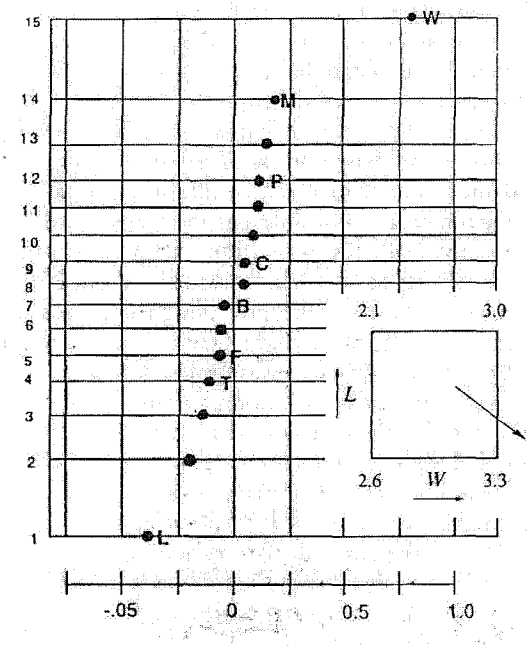


Figure 5.

a) *Construction of a paper helicopter.*

b) *Results from 16 run fractional factorial experiments showing the factor levels and the calculated main effects of the eight factors.*

c) *A normal plot of the effects from the helicopter experiment. The inset diagram summarizes the conclusions.*

made yesterday when it was cold and wet, the runs with the blue helicopter were made today when it's warm and dry. Perhaps it's the temperature or the humidity that made the difference. What about the paper? Was it the same kind of paper used to make the red helicopter as was used to make the blue one? Also, the blue helicopter was dropped by Tom and the red one by Dick. Perhaps they don't drop them the same way. And *where* did Dick drop his helicopter? I bet it was in the conference room, and I've noticed that in that particular room there is a draft which tends to make them fall towards the door. That could increase the flight time. Anyway, are you sure they dropped them from the same height?". So we ask the class if they think these criticisms have merit and they mostly agree that they have, and they add a few more criticisms of their own.

We tell the class how such considerations led Fisher some seventy years ago to consider the precautions necessary in running an experiment so that it can provide data which leads to unambiguous conclusions rather than to an argument. We show how his ideas can be used to compare the blue and the red helicopter by making a series of paired comparisons. Each pair (*block*) of experiments involves the dropping of the blue and the red helicopter by the same person at the same location. The decision as to which helicopter should be dropped first is made *randomly* by tossing a penny. The conclusions are based on the *differences* in flight time within the pairs of runs made under identical conditions. We go on to explain however that different people and different locations could be used from *pair to pair* and how, if this were done it would strengthen the extrapolability or as Fisher (1935) said "widen the inductive basis" of the experimental results. If the red helicopter design appeared to be better, one would, for example, like to be able to say that it seemed to be better no matter who dropped it or where it was dropped.

A fractional factorial design. As another example of the use of this device; at a later stage in the course it is used to illustrate the value of fractional factorial designs as screening devices. It is supposed that a brainstorming session by an engineering design team has led to the selection of eight factors to be studied. These selected factors are listed at the top of Fig. 5(b) together with the two conditions (indicated by minus and plus signs) at which each will be tested. It is thought likely that only a few of these factors will have important large effects. We are thus in the familiar "Pareto" situation where, as Dr. Juran says, we want to screen out "the vital few from the trivial many". The design used in Fig. 5(b) is a 2^{8-4}_{IV} fractional factorial. The student may be taught something of the theory of these designs; however, to use them, all they really need is a table such as has been supplied by Bisgaard (1988) which gives this and other eight and sixteen-run designs with a succinct description of their properties and

analysis. As is well known, the 2^{8-4}_{IV} design has two very valuable characteristics:

a) if there are interactions between pairs of factors they will not bias any of the 8 main effects of the factors;

b) if only up to 3 factors are of importance, the design will produce a complete 2^3 factorial design replicated twice in those three important factors no matter which ones they are (see, for example, Box, Hunter and Hunter, 1978).

Flight times for the sixteen helicopter types obtained from an experiment run in random order are also shown in Fig. 5(b). From these flight times, 8 main effects and 7 strings of two-factor interaction may be calculated on the assumption that interactions between 3 or more factors may be ignored. The effects are plotted on probability paper in Fig. 5(c) suggesting that real effects are associated with W (wing length) and, less certainly, L (body length). On the basis that the remaining effects falling around the straight line are mostly due to noise, we can summarize the data simply in terms of the inset diagram in Fig. 5(c). The experiment immediately provides not only an improved helicopter design but also indicates the direction in which further experimentation should be carried thus initiating a sequential process of experimentation which can be carried as far as one desires.

Another aspect of this approach is highlighted by discussing with the class whether they are satisfied with *flight time* as the sole criterion. In earlier lectures we have emphasized to the class that what happens in each run of an experiment must be carefully documented - for example the fact that helicopter #7 hit the table leg and that the run had to be repeated. The need for careful observation is emphasized, perhaps leading to the conclusion that an additional criterion such as flight *stability* should be considered in future experimentation. This teaches the lesson that appropriate and feasible objectives *cannot* always be determined in advance. The *criteria* to be used in assessing the results may need to be modified or totally changed during an investigation as more is learned. The helicopter can also be used to illustrate the process of iterative experimentation at later stages. For example (a) to demonstrate the use of experimental design to reduce variance, (b) to illustrate the importance of estimating variance components and of using the appropriate error term and (c) to consider the problems of obtaining a robust product or process.

Acknowledgment

This work was supported by a grant from the Alfred P. Sloan Foundation.

Bibliography

Bates D.M. and Watts D.G. (1988), *Nonlinear Regression Analysis and Its Application,* John Wiley & Sons, New York.

Bisgaard S. (1988), *A Pratical Aid for Experimenters,* Starlight Press, Madison, Wisconsin, USA.

Bisgaard S. (1991), Teaching Statistics to Engineers, *The American Statistician,* 45(4), pp. 274-282.

Box G. (1990), Must We Randomize Our Experiment, *Quality Engineering,* 2(4), pp. 497-502.

Box G. (1992), Teaching Engineering Experimental Design with a Paper Helicopter, Quality Engineering, 4(3), pp. 453-459.

Box G. (1993), Statistics and Quality Improvement, *Journal of the Royal Statistical Society,* Series A, 156, pp. 209-229.

Box G., Bisgaard S. and Fung C. (1988), An Explanation and Critique of Taguchi's Contributions to Quality Engineering, *Quality and Reliability Engineering International,* 4, pp. 123-131.

Box G. and Draper N. (1987), *Empirical Model-Building and Response Surfaces,* John Wiley & Sons, New York.

Box G.E.P., Hunter W.G. and Hunter J.S. (1978), *Statistics for Experimenters,* John Wiley & Sons, New York.

Box G. and Kramer T. (1992), Statistical Process Monitoring and Feedback Adjustment-A Discussion, *Technometrics,* 34(3), pp. 251-285.

Box G. and Newbold P. (1971), Some Comments on a Paper by Coen, Gomme and Kendall, *Journal of the Royal Statistical Society,* Series A, 134, pp. 229-240.

Box G.E.P. and Wilson P. (1951), On the Experimental Attainment of Optimum Conditions, *Journal of the Royal Statistical Society,* Series B, 23, pp. 1-45.

Deming W.E. (1950), *Some Theory of Sampling,* John Wiley & Sons, New York.

Deming W.E. (1986), *Out of the Crisis,* MIT Press, Cambridge, MA.

Finney D J. (1945), Fractional Replication of Factorial Arrangements, *Annals of Eugenics,* 12, pp. 291-301.

Fisher R.A. (1935), *The Design of Experiments,* Oliver and Boyd, Edinburgh and London.

Hunter W.G. (1977), Some Ideas About Teaching Design of Experiments, with 2^5 Examples of Experiments Conducted by Students, *The American Statistician,* 31(1), pp. 12-16.

Ishikawa K. (1976), *Guide to Quality Control,* Asian Productivity Organization, Tokyo. Available in USA from UNIPUB, New York.

Michaels S.E. (1964), The Usefulness of Experimental Design (with discussion), *Applied Statistics*, 13(3), pp. 221-235.

Morrison S.J. (1957), The Study of Variability in Engineering Design, *Applied Statistics*, 6(2), pp. 133-138.

Plackett R.L. and Burman J.P. (1946), The Design of Optimum Multifactorial Experiments, *Biometrika*, 33, pp. 305-325.

Taguchi G. (1986), *Introduction to Quality Engineering*, White Plains, New York.

Tukey J.W. (1977), *Exploratory Data Analysis*, Addison-Wesley, Reading, MA.

Youden W.J. (1961a), Experimental Design and ASTM Committee, *Materials Research and Standards*, 1, pp. 862-867. Reprinted in H. H. Ku (ed.), *Precision Measurement and Calibration*, Vol. 1, Special Publication 300, Gaithersburg, MD, National Bureau of Standards, 1969.

Youden W. J. (1961b), Physical Measurement and Experimental Design, *Colloques Internationaux de Centre national de la Recherche Scientifique* No. 110, le Plan d'Expériences, pp. 15-128. Reprinted in H. H. Ku (ed.), *Precision Measurement and Calibration*, Vol. 1, Special Publication 300, Gaithersburg, MD, National Bureau of Standards, 1969.