CUTTING EDGES IN BIOMETRY*

by

Walter T. Federer
337 Warren Hall
Cornell University
Ithaca, New York, 14853

BU-809-M

Rev. October, 1983

* Biometric Society, ENAR, President's Invited Address, Nashville, Tennessee, March 22, 1983.

Cutting Edges in Biometry*

Walter T. Federer

Biometrics Unit, Cornell University, Ithaca, NY 14853, U.S.A.

SUMMARY

Our purpose here is to demonstrate that cutting edges exist in all areas of biometry. After making comparisons of papers in eight volumes of Biometrics over a 31-year period and of those presented at the last International Biometric Conference, Toulouse, France (September 1982), cutting edges in several areas are discussed. Several areas of research on biometrical methods for problems of plant and animal agriculture are explored, e.g., the statistical design and analysis of experiments on intercropping (the simultaneous growing of several crops on the same unit), the transfer-of-information, pasture experiments, toxicological studies, sampling plans for environmental pollution, and difficulties with repeated measures design. Another group of topics discussed are calibration, measurement, and quantification of responses. These problems are often ignored. Four specific situations are discussed. A third topic discussed pertains to laboratory analyses. A laboratory overload can be alleviated by using statistical procedures such as subsampling, pooling subsamples, group testing, sequential sampling, double sampling, and obtaining running estimates of analytic

---

* President's Invited Address, Biometric Society (ENAR), Nashville, Tennessee, March 22, 1983.

Key words: Intercropping; Laboratory analyses; Environmental pollution; Calibration and measurement; Group testing; Sequential sampling; Pooling subsamples; Model construction and selection; Biometrical research areas; Repeated measures design.

errors with a small number of samples. A fourth area presented is model construction and selection. Some examples are cited, and the problem of how random errors arise is discussed. A number of other areas – assays, pattern analysis, and standardization – are discussed briefly. The paper ends with an encouragement for statisticians to research these problems.

1. Introduction

To eliminate any misunderstanding, we start with definitions for terms that may have several meanings. As biometry can, and does, have various interpretations, and as Webster's definition of it, "the statistical study of biological observations and phenomena," is too confining and too restrictive, the following definition is used by the author:

> Biometry is the study, development, and application of procedures and techniques in computer science, mathematics, operations research, probability, statistics, and systems analysis for biological investigations and phenomena.

Some other useful definitions are:

Biometrician – a specialist in biometry,

Biometric – an adjective whereas biometrics is a noun synonymous with biometry,

Biostatistics is Webster's definition of biometry, but it should not be a synonym for medical statistics, as used in some quarters,

Biomathematics – the study, development, and application of mathematics for biological investigations and phenomena. (It is not a synonym for biometry, although some individuals who consider statistics to be solely in "the mathematical sciences" may argue otherwise.)

These items received considerable discussion at the Symposium on the Development and Implementation of Courses and Curricula in Natural Resources-Biometry (National Science Foundation and Colorado State University, April 20-24, 1970, W. E. Frayer, Symposium Director and Editor of the Proceedings).

As a base, articles in eight volumes of Biometrics spanning a 31-year period and the papers presented at the latest International Biometric Conference were classified into nine categories. This was done to try to find out how the cutting edges of Biometry were being attacked. Each category is discussed further by pointing out needed areas of research. Answers to some of the questions are needed for further and faster progress in science. Many problems are cited, and there are many more. Also, it should be noted that biometrical and statistical articles are appearing in several other journals, notably those in biology and agriculture.

2. Types of Articles in Biometrics and Papers at Toulouse

To ascertain the types of papers published in Biometrics the papers published in the years 1950-51, 1960-61, 1970-71, and 1980-81 were classified as follows:

Agriculture – Statistical procedures developed for and/or applied to plant or animal experiments.

Animal populations – Statistical procedures developed for and/or applied to fish, fowl, and animal population ecological studies.

Assays – Statistical procedures developed for and/or applied to biological assays.

Biological – Statistical procedures developed for and/or applied to all types of biological investigations. The classification "animal populations" could have been included here, as there is considerable overlap. This classification is more plant and ecologically oriented than animal populations.

Clinical trials – Statistical procedures developed for and/or applied to investigations involving clinical studies. This categorization is somewhat broader than confining attention to randomized trials to compare the effectiveness of medical treatments on patients.

Medical – Several statistical procedures have been developed for and/or applied to investigations on medical aspects above and beyond clinical trials which could have been included under this category.

Genetics and breeding – Statistical procedures developed for and/or applied to investigations concerning genetics and/or breeding; they have a long history, and research remains active in this area at the present time.

Modeling – Much of past and present statistical procedures have assumed a model, and "general linear model" theory is based on the assumed response model. There have been occasional studies involving the determination of the actual response model, rather than assuming one. There has been a small amount of research on developing and applying response models to data from investigations over the years; there appears to be an increasing awareness of this important aspect of statistical investigation.

Statistical and biometrical methodology – Statistical and/or biometrical research with no specific applications, e.g., tests of hypotheses, construction of orthogonal latin squares, variance component estimation, computer programs, etc.

Any classification such as the above is arbitrary and subjective. Given a description of a classification, one needs to be certain that there was consistency in grouping the papers over the items being classified. Care was taken here to assure that the groups were such that the groupings of papers were repeatable when a reclassification of the same articles was made. The groupings were selected to ascertain the changes that might occur over time. The numbers of papers in the nine groups for eight volumes of Biometrics are given in Table 1.

Insert Table 1 here

The main upward trends in Table 1 are for medical statistics and for clinical trials. The main downward trend is for statistical and biometrical papers with no area of application. Papers on assays are down somewhat from the 1950-51 period, while those on biological statistics are increasing slightly over the years. Genetics and breeding and modeling papers are somewhat higher in the 1960-61 and 1970-71 volumes than in the other two periods.

In addition to the papers in the above eight volumes of Biometrics, the papers presented at the XIth International Biometric Conference in Toulouse, France, September 6-10, 1982, were classified likewise in Table 2. A comparison of percentages of the Toulouse papers with the 1980-81 volumes of Biometrics indicates an increase for agricultural statistics, biological statistics, modeling, and statistics and biometry papers.

Insert Table 2 here

There were 19% of the papers at the Toulouse meetings on medical statistics, whereas there were more than double this percentage, 40%, published in the 1980-81 volumes of Biometrics. In order to determine which cells in Table 1 were giving large deviations from a chi-square contingency computed value, the cell contributions for a contingency table chi-square value are given in Table 3. Clinical trials gave large contributions, since it is only in recent years that papers on this topic have been published in Biometrics. Medical statistics also gave large contributions to the total chi-square value of 134. Statistics and Biometry papers in the 1970-71 and 1980-81 volumes were much fewer than one would expect from the computed values.

Insert Table 3 here

To determine if the totals in the last column of Tables 1 and 2 were in the same relative proportions, a contingency chi-square was computed. The contributions for the individual cells are given in Table 4. Large row contributions to the total chi-square value of 45 came from agriculture, animal populations, assays, clinical trials, genetics and breeding, and modeling. For both sets of papers considerable heterogeneity was encountered.

Insert Table 4 here

3. Some Cutting Edges in Agricultural Statistics

Articles on aspects of agricultural statistics have been few and far between in the eight volumes of Biometrics (1%). This topic received somewhat more attention at the International Biometric Conferences in that 6% of the papers were on this topic in the last one. An age-old practice in plant agriculture is intercropping, the growing of two or more crops simultaneously or successively on the same plot of land. Within the last 20 years this topic has received attention from agricultural researchers and a few statisticians (see, e.g., Mead and Riley, 1981). The statistical problems of design, analysis, and inference associated with experiments on intercropping are many and varied. One basic problem is how to combine responses from all the crops grown on a plot of land for economic returns, disease control, insect control, fertilizer replacement, stability, stability versus maximum yield, calorie yield, protein yield, starch yield, optimization of land and labor usage in light of goals, and comparisons of farming systems.

Another basic problem in intercropping research is how to study spatial arrangements, density relationships, and orientation of plants and crops. The number of possible arrangements and relationships is quite large. What is the

best way to reduce this number and obtain the desired information? With any scheme that one comes up with, how does one check whether the results are useful in practice?

A third large problem in this area is the development of appropriate response model equations and competition models that adequately describe responses for a given crop mixture. How does a changing proportion of crops affect the response and competition models? How are effects in a model to be defined with changing proportions? An answer to these questions will necessarily have to be made in collaboration with the agricultural researcher, with the statistician as an active collaborator.

Statistical design problems will be many and varied, varying from describing the population, sampling unit, experimental unit, and response model to treatment design (selection of treatments) and experiment design (arrangement of treatments in an experiment) to determination of replicate number and allocation, experimental unit size and shape, elimination of competition between experimental units, and to determination of how to measure response.

Some specific questions that require answers are:

(i) How does one measure stability once it has been defined?

(ii) What is the statistical distribution of the stability measure? Should it be tabled for various values of the parameters?

(iii) How does one extend present multivariate analysis theory to allow comparison of sole crops (one crop grown on a plot of land) and mixtures of crops, allow for heterogeneous variances and covariances, allow for comparisons of non-normally distributed variates, and determine just how valuable presently available multivariate procedures are for combining responses?

(iv) What are the properties of various statistical methods for combining responses such as a relative land use measure (land equivalent ratio), a relative

economic index, calorie index, or a protein index? What are their distributional properties? Which ones should be tabled for mixtures of 2, 3, ..., v crops?

(v) How does one evaluate intercropping systems over time?

These and other questions will require the efforts of many creative statisticians with a willingness to leave well-formulated problems such as those associated with linear model theory and hypothesis testing. The one-sample, one-response, one-population, and/or i.i.d. world is not tenable in this area.

Relay, sequential, and rotational cropping (the growing of crops in sequence) experiments will be faced with many of the difficulties mentioned above. Results obtained for the above may be useful here, but new procedures will be necessary. It is amazing how little follow-up has been done for the research efforts of Yates (1949) and Cochran (1939) in the 30's and 40's on rotation experiments. They did not solve all the problems, as was demonstrated by Patterson (1964). The complexity and innovativeness required appears to have frightened statisticians away from this area. These results are necessary in evaluating farming systems in present-day investigations.

Another large and provocative area for research in agriculture is the how, why, where, and what of transferring information from a greenhouse, a growth chamber, or a field experiment into valid recommendations for a farmer. How far can one go with any of the available procedures? How can one use present procedures, or new ones, to transfer experimental information into practice to the same or to other regions?

A research area encompassing both plants and animals is pasture experiments. A manual on procedures, design, analysis, and inferences is much needed and past due. Many pasture experiments are conducted, but the results have questionable design, analysis, and inference aspects. What should be considered in "stocking rate" (the number of animals required to fully utilize a pasture), "carrying

capacity" (the number of animals a pasture would sustain over a season), long-term effects of a grazing system, the statistical design and analysis, and other problems in pasture experiments?

The sizes of dams in the beef cattle industry vary widely, as do the recommendations for size. How does one estimate the most economical size of dam for rearing baby beefs for any particular breed and environment, or for all cattle in general? How does one go about constructing criteria for optimum size of dam and how does one determine optimum size?

In toxicological studies on animals or humans, how does one determine when an effect becomes "toxic"? For various criteria how valid are present U.S. Food and Drug Administration (FDA) numbers of animals (dogs, rats, monkeys, cows, pigs)? How should standards be set? How does one design and analyze for possible unknown side-effects of drugs, diets, handling procedures, etc.? What is an appropriate response model in a toxicological study, and how is food intake to be taken into account? In this same connection, if a response reaches a plateau or high point for various amounts or doses of a given item, how does one design and analyze data to estimate the lowest effective level or dose? How does one determine an "optimal" design for such situations? If an "optimal" design is not possible, what are "good" designs?

What is an optimal design for sampling from toxic chemical dumps, sudden releases of radiation, garbage dumps in the ocean by large cities, or other environmental pollution? Should the samples be taken in concentric circles, ellipses, or some other contour arrangement around the site? What should be the frequency of sampling on each of the contours? How frequently should the site be sampled? What characteristics should be measured to determine damage or suspected damage to the environment? In legal suits, what data are necessary to substantiate or disprove cause of damage?

In connection with using repeated measures designs on plants, animals or humans, there are many unsolved and undiscussed problems. First of all, it would be highly desirable in designing repeated measures experiments to have some idea of the response curve for a treatment plotted against time. The length of the treatment period should be long enough for the treatment effect to be asserted. Correct choice of a treatment period may eliminate carry-over effects or perhaps limit them to one additional period. In a repeated measures design, is it advisable to have sampling units which receive the same treatment for p periods? In what sense would a "good design" include sequences wherein a treatment followed itself for p - 1 successive periods? Under what situations would a "good" design include pre- and/or post-treatment periods? Under what situations would a design with one pre-treatment period and one treatment period be "better" than a two-period treatment design?

What types of experimental material are amenable to the use of repeated measures designs? Suppose that a response is easy to decrease but very difficult or impossible to increase at any given time. To illustrate, milk yield in dairy cows is very easy to reduce in the latter part of their lactations, but it is very difficult to increase the level. How should treatments be applied in a repeated measures design to account for this? What is an appropriate statistical analysis for such situations?

4. Cutting Edges in Calibration, Measurement, and Quantification

Many measuring instruments are not finely enough calibrated to measure the desired quantities. Thermocouples may not be sufficiently calibrated to measure differences in temperature due to respiration of seeds. Chemical analytic procedures may not be suitable to measure minute amounts of a chemical substance. Can this problem be overcome by changing the technique and/or using appropriate statistical

designs and analyses? To illustrate, suppose that one has a table of unknown length and that one has a measuring instrument accurately calibrated in feet. Without recalibrating the instrument to units smaller than one foot (which might be the correct thing to do), how does one measure the length of the table? First, one must define what is meant by the length of the table, as the shape was not given. Once this has been determined, the measuring stick is turned on the table along the line of its length until it comes out exactly on a one-foot mark (or even at the end of the measuring instrument). The length of the table, L, is given by the length of the stick, L^*, times the number of lengths of measuring instrument, c, divided by the number of turns of the measuring instrument, n, or $L = L^* c/n$.

Measurements are sometimes categorized into classes; how does one quantify the distance between classes? To illustrate, dormancy in plants may be categorized into 11 classes, 0,...,10, with zero meaning 0% dormant and 10 meaning 100% dormant; then, another class, dead, is added as class 11, to make 12 classes. Now, is death one, 100, 1000, ..., units away from plants that are 100% dormant? It could be construed as infinite in some cases. One can obtain any level of significance desired, if there are dead plants in the experiment, by assignment of a value to the dead class. Also, suppose that one treatment kills most, or all, of the plants, but the remaining treatments break down dormancy to a certain degree. How does one include these results in significance testing and interval estimation?

In certain types of experiments, measurements go from zero to some upper bound with the upper bound decreasing as the level of a treatment increases. Suppose it is desired to estimate the line going through the mean of a level plus two standard deviation units, which estimates the maximum number in some

sense for each level. An example where this occurs is in weed control by mulching. For no mulch, the number of weeds per experimental unit can go from zero to some large number, say N_0. As mulch is applied the maximum number of weeds, N_i for level i, decreases as i increases. N_i becomes zero for i large. How does one estimate the effectiveness of level i in controlling weeds? How does one estimate an upper bound on the weed population for level i with a specified level of probability?

Many types of data include values of zero, of a trace, of a small amount, and then a series of numerical values. How does one go about analyzing such data? How far is zero from a trace, and how far is a small amount from a trace? In some sets, a fairly large proportion of the data is graded as a trace. How does one quantify such data and handle the quantified data in a statistical analysis?

5. Cutting Edges in Laboratory Analyses

Analytic laboratories for soil, plant, animal, human, and other material are frequently overloaded. The laboratory director's solution is to hire more people and have a larger facility and organization. What is the statistician's solution? Most of them have none, because they forget about the basic element of statistics; that is, something called sampling. True, the material from one investigation is a sample in itself, but we can do various things to this sample. We can

(i) subsample the samples,

(ii) pool samples from the sample,

(iii) sequentially sample from the samples with some acceptable stopping rule,

(iv) use double sampling procedures when two methods of analysis are available, with one being quicker and cheaper, or

(v) subsample some or all individual samples for checking on analytic variation (this increases the number of laboratory determinations).

Because of the overload and priorities, many samples never get taken, or if they are, they may never be analyzed. This problem could be eliminated, or at least alleviated, if the statistician becomes involved and if sampling is used. A thorough statistical and subject matter study is needed to determine methods of reducing the number of analyses, of obtaining quicker and easier analytic methods, methods of stratifying samples to isolate major sources of variation, and other procedures which minimize the time between sampling and having samples analyzed and which reduce the cost of analyses.

When an analysis for a certain characteristic is difficult or impossible to obtain because of a laboratory overload or because of cost, an experimenter needs to determine if analyses from a subsample of the samples will provide satisfactory information. If not, then the analyses will not be done, and the consequences of this action need to be assessed. Alternatively, the experimenter should determine if the goals of an experiment can be achieved by pooling the samples. For example, since iodine number in soybeans varies little from block to block in a field experiment, the samples from the r replicates for one treatment are pooled, and an iodine number is obtained for the pooled sample. This reduces the number of laboratory analyses to 1/r of that for individual experimental units. In determining arsenic levels in human or animal hair, where locality or region differences are the only important items, it is reasonable to pool equal amounts of hair from each of the types of individuals in a locality. The same could be true for selenium levels in human blood samples. For example, in an impending survey, 7200 individual blood samples are to be taken. Since the present goal is for selenium levels in young (under 30) and old (over 30), by sex, and for 72 localities, analyses will be done on $2 \times 2 \times 72 = 288$ samples instead of 7200. This involves only $1/25 = 4\%$ of the analyses required by doing individual analyses. This was well within the laboratory's capabilities, and

many more characteristics could be obtained. Another recent example was in a soil compaction experiment where four trees were grown per pot. Instead of doing an analysis on each tree, analyses were done on samples from a composite of the four trees. This procedure was within the experimenter's capabilities, whereas doing individual trees was not; little additional information would have been acquired by doing individual tree analyses.

Group testing is an idea that has been in statistical literature since 1946, but almost all directors of analytic laboratories, and perhaps most statisticians, are unaware of its existence and possible usefulness. Its usefulness will be to determine presence or absence, or less than a given amount of a substance, in the pooled sample. The original use was to reduce the number of analyses for detecting presence or absence of syphilis in army recruits during World War II. This became known as the Dorfman group testing procedure for rare occurrences. Some work has been done on optimal group testing procedures (e.g., Sobel and Groll, 1966, and Bush et al., 1980, "New combinatorial designs and their applications to group testing," Technical Report No. BU-726-M in the Biometrics Unit series, Cornell University, Ithaca, NY) and on optimal group size using statistical criteria. This is insufficient as the dilution rates that are detectable analytically must be taken into account. What needs to be done here is

(i) to determine how rare is rare, i.e, what percentage of samples have the characteristics to make group testing more efficient than individual sampling,

(ii) to determine which characteristics are amenable to group testing in any given laboratory,

(iii) to determine the lowest dilution rates that can be detected using present analytic procedures,

(iv) to determine the optimal group size using statistical criteria and then use this number or that in (iii), whichever is smaller, and

(v) using the group size in (iv), to determine an optimal procedure for group testing.

Double sampling procedures will be useful when there are two procedures for analyzing samples, one a quick and/or cheap method and the second slower and/or more costly. The first method is used on all samples, and the second on only a small fraction of the samples. Amount of information per unit of cost (or time) needs to be ascertained for each characteristic and pair of analytic methods. A high correlation between results of the two methods is required for double sampling to be viable.

Sequential sampling procedures will be useful when samples are taken and analyzed sequentially or when all samples must be taken at a particular time, but they are analyzed sequentially and the cost of analysis per sample is not negligible. To illustrate, in research on sheep, 50 sheep had been allotted for an experiment with destructive sampling. When six sheep had been slaughtered, the results became obvious. When 12 sheep had been slaughtered, the experimenter, in desperation, called a statistician requesting permission to stop and asking if he did stop how could he justify the procedure. He did not want to have to justify a result guided stopping procedure, and he did not want to slaughter the remaining 38 sheep needlessly, as this was very costly. He was told to take three more sheep and stop, to make the stopping rule almost free of a criticism of using a result guided one. But, really, shouldn't he have stopped at six rather than 15? How good was the 15-sheep stopping rule? What procedure should the experimenter have used?

As a second example where sequential sampling may be used efficiently, consider a toxic chemical dump, e.g., Love Canal (which is located near Buffalo, New York), where it is desired to know if the chemical dump affects animals or humans, or has polluted the soil and/or air at varying distances from the dump. Suppose that the sampling plan was to set up concentric circles around the dump site, and to take equally spaced samples on each of the concentric circles.

Further, suppose that the samples all had to be taken because of political and other pressures, and that the cost of analyzing the samples was relatively large, say $2000 per sample. Sequential analysis of the samples is definitely indicated here. One would first analyze the samples on the inner ring and a very small percentage of outer ring samples, which could be used to detect sample deterioration over time. Then one would do the second inner ring out and a small percentage from the other rings, etc. Once one reaches a ring where there is no contamination, one stops, unless contamination has shown up in the analyses of the small percentage of samples from the outer rings. Proceeding in this fashion one may need to analyze only a small fraction of the total samples. This is what could have been the result of the Love Canal sampling. A savings of millions of dollars could have been achieved. (In the Love Canal case, outside pressure may have made it necessary to analyze all samples, regardless of the outcome on the first samplings in the sequential plan.)

Turning to the other side, i.e., increasing analyses, and doing duplicate, triplicate, etc., analyses on a sample, one should realize that duplicate analyses on n samples result in 2n analyses, triplicate analyses on a sample result in 3n analyses, etc. This is often done in the name of checking on analytical variation. Usually 10-20 duplicate observations would suffice to give the desired estimate of analytic variation (with 10-20 degrees of freedom) to compare with sampling variation. It is suggested that 10-15 duplicates be included in any sampling procedure to check on analytic error. In the selenium example, 12 observations on a known sample were included with the 288 samples as an analytic check and to obtain an estimate of the analytic error variance and bias. The 300 samples are done in a random order. Duplicate determinations or readings double the amount of observations, and thereby increase the cost of the statistical analyses. If the relative amount of information is small, the extra cost should not be allowed. Too often, no thought is given to this aspect, resulting

in inefficient experimentation. Obtaining estimates of variation from the various possible sources of variation can be accomplished by allocating 10 to 20 degrees of freedom for this. Once one locates an important source of variation, additional resources would be allocated to take this into account. It is suggested that a continuing check be made on possible sources of variation with small allocations of degrees of freedom, say 5 to 10, through time.

One additional question here is: When and how will quality control procedures be used on experiments, surveys, and other investigations to determine their suitability for publication? Experiments, surveys, and investigations differ widely in quality but are usually treated as equal when published.

6. Cutting Edges in Model Selection

The author predicted in the 1950's that model construction and model selection would become one of the more important aspects of statistics within the next 10-15 years and would occupy the attention of a large proportion of statisticians. One good paper (Box and Cox, 1964) has appeared. Even if this prediction was wrong this should happen by the turn of the century; 21st century statisticians will most likely be occupied with this concept. The development of "modern data analysis" is a tool for aiding in the selection of models, but more tools are required. Methods for detecting outlying treatments and outlying blocks are required. Multivariate outlier theory needs considerable extension. Also, other techniques are needed to obtain appropriate response model equations.

Perhaps the biggest obstruction in the area of modeling has been the statistician's obsession with Taylor series expansions in terms of the polynomial regression model and the linear model. At best, the latter is a linear model, and usually a poor approximation of the true underlying response model in a real world situation. In many, if not most, areas of investigation, a nonlinear form

of a response model would be more appropriate than the linear one used. The error deviation has been assumed to be an add-on, and statisticians are prone to defining an error vector as $\underline{\epsilon} = \underline{Y} - E\underline{Y} = \underline{Y} - X\underline{\beta}$, where $\underline{Y}$ is an observation vector and X is a design matrix. Now how much of $\underline{\epsilon}$ is part of the stochastic process and how much is an add-on? In field experiments is there an error parameter ϵ_{ij} attached to the ijth experimental unit? Is it constant regardless of which treatment lands on that experimental unit? Or, is ϵ_{ij} mostly a treatment x experimental unit interaction component? If the latter, what good are permutation tests?

In agriculture, modeling of crop, plant, and animal responses is required for more rapid advance of theory in these fields. Forms of response equations are needed when one, two, or more elements are limiting for plant and animal nutrition experiments. Knowledge of the minimum and the maximum levels of n - 1 factors which allow full expression of an nth factor at increasing levels would be desirable. Limiting levels of any of the n - 1 factors for limiting any given level of the nth factor would be desirable. How does one estimate minimal effective level or dosage, conditional upon given levels of the other n - 1 factors? How should recommendations be made for general public usage?

To date, work on model construction and selection has been sporadic and uncoordinated. Several papers are concerned with whether or not one should include quadric or cubic terms in a polynomial regression model. Given that the polynomial model may not be an appropriate one, the results obtained are academic and of little practical usefulness. Modeling has been receiving more attention at statistics meetings, especially on the international level, than in journals. More sessions on model construction and selection are needed in order to develop a literature on the principles of modeling responses in agriculture, biology, and medicine. Perhaps the entire theme of several future statistics and/or a subject matter area meetings will be modeling. The development of computer programs such

as GLIM is a help but more general programs, especially for nonlinear models, are needed.

Deterministic modeling may have a place in the area of stochastic modeling when errors are simply additive to a deterministically regulated process. However, when the error terms are a part of the process, then deterministic modeling may have little or no usefulness.

7. Cutting Edges in Other Areas

A number of other areas requiring considerable research effort to meet existing and future problems arising in the statistical design and analysis of experiments are listed below. This is, by no means, a complete list.

Assays: Since assays are an integral part of biological experimentation, and since most introductory statistics books do not mention this important topic, how should this omission be remedied? Since biologists use ratios and statisticians use differences on the same scale, how should this be remedied?

Pattern analysis: Much biological, plant, animal, and meteorological data come in the form of a frequency polygon, i.e., in graphical form. Some examples are the flight of a missile, the flight of a homing pigeon, temperature over a 24-hour period, an electrophoretic graph for a blood sample, etc. Since each graph is one datum, how should these data be analyzed statistically? If one has 15,000 such graphs, how does one summarize the data? If one knows the mathematical function describing the graph, one could obtain estimates of the parameters for each graph and perform a multivariate analysis on the estimated parameters. What are appropriate statistical methods of analysis? What happens when there are errors in the abscissa of the graphs?

Standardization: Many laboratories perform the same type of analysis on samples from plants, animals, humans, soils, air, water, etc. Even though n

laboratories could have identical samples, there is often no assurance that the same results would be obtained. In fact, it is quite likely that different results would be obtained. One reason is that the procedure, the equipment and/or the technique may not have been standardized. If scales, thermometers, etc. are not standardized, one will usually not obtain the same readings. Sometimes a procedure has not been appropriately documented to allow everyone to use it in the same manner. Many assays and analytical procedures are in the mind of the experimenter, who simply has never bothered to write up the procedures in sufficient detail for use by other people. Some form of quality control would be highly desirable. The statistician will be useful in devising such procedures, helping to establish controls, in determining limits of allowable error, etc.

In other areas, state and federal regulations require certain sample sizes and other requirements. For example, one noxious weed seed in a five-gram sample may be unacceptable, whereas $\frac{1}{2}$ gram of non-noxious weed seeds may be allowable. In toxicity studies, the FDA may require three monkeys, six dogs, and 20 rats of each sex. On statistical grounds, why should the required numbers differ? Why aren't they the same? How does one ascertain that there is not one noxious weed seed in a lot of seed?

Others: Such areas as medicine, clinical trials, animal populations, ecology, genetics, etc. require many new and innovative statistical and biometrical procedures to satisfy their needs. Forcing these needs into presently available biometrical technology may be completely unsatisfactory, and perhaps misleading.

8. Epilogue

As demonstrated above, the cutting edges of biometry are many and varied. Also, as demonstrated, many of the cutting edges have been ignored, considered unimportant, or otherwise relegated to someone else. Research in many subject areas

has reached a stage of development where answers to some of the above are required. Statistics has not kept pace with the needs and has become more narrowly focused in some areas. Hypothesis testing in the Neyman-Pearson lemma spirit, combinatorics of experiment designs, linear model theory development, so-called distribution-free (nonparametric) procedures, etc. have tended to distract statisticians' attention from real life problems. One reason is that many statisticians are concerned with mathematical niceties such as writing lemmas and theorems; many of the problems discussed here require little or no mathematics, but may use computer simulations for some statistics and analyses of large data sets for others. Since these cannot be mathematized easily, statistics journals may be reluctant to publish these research results. Subject matter journals may be the better place to publish such material.

The real world of statistical applications is challenging, interesting, and rewarding. One will not be able to matricize or integralize one's way to answering a question, but there are other ways! Statisticians can make major contributions in the various subject matter fields by solving their important statistical problems. They should consider how much of what has been published in Statistics during the past 50 years is useful in subject matter research areas today. For example, how useful are decision theory and hypothesis testing for present-day experimenters?

References

Box, G. E. P. and Cox, D. R. (1964). An anlysis of transformations (with Discussion). Journal of the Royal Statistical Society, Series B 26, 211-252.

Cochran, W. G. (1939). Long-term agricultural experiments. Journal of the Royal Statistical Society, Series B 6, 104-148.

Mead, R. and Riley, J. (1981). A review of statistical ideas relevant to intercropping research (with Discussion). Journal of the Royal Statistical Society, Series A 144, 462-509.

Patterson, H. D. (1964). Theory of cyclic rotation experiments (with Discussion). Journal of the Royal Statistical Society, Series B 26, 1-45.

Sobel, M. and Groll, P. A. (1966). Binomial group testing with an unknown proportion of defectives. Technometrics 8, 631-656.

Yates, F. (1949). The design of rotation experiments. Commonwealth Bureau of Soil Science Technical Communication Number 46 (Proc. First Commonwealth Conf. on Tropical and Sub-Tropical Soils, 1948), pp. 142-155. Harpenden, England: Commonwealth Bureau of Soil Science.

Table 1

Classification of papers in Biometrics for selected periods

Classification*	Period 1950-51 No.	%	1960-61 No.	%	1970-71 No.	%	1980-81 No.	%	Total No.	%
Agriculture	2	3	2	2	0	0	3	2	7	1
Animal Populations	3	4	6	6	10	6	11	8	30	6
Assays	8	12	6	6	4	2	10	7	28	6
Biological	9	13	14	14	23	14	9	7	55	12
Clinical Trials	0	0	0	0	0	0	11	8	11	2
Genetics and Breeding	7	10	16	16	25	15	16	12	64	14
Medical	9	13	4	4	20	12	54	40	87	19
Modeling	2	3	13	13	14	8	9	7	38	8
Stat. and Biom. Methods	27	40	39	39	70	42	11	8	147	31
Total (% within rounding error)	67	100	100	100	166	100	134	100	467	100

* Any classification of this type is subjective, even the categories selected.

Table 2

Classification of papers presented at the

XIth International Biometric Conference, 9/82

Classification	Invited		Contributed		Total	
	No.	%	No.	%	No.	%
Agriculture – plant	2	6	12	5	14	5
animal	0	0	4	2	4	1
Animal Populations	1	3	4	2	5	2
Assays	2	6	5	2	7	3
Biological	1	3	36	15	37	14
Clinical Trials	2	6	12	5	14	5
Genetics and Breeding	4	12	15	6	19	7
Medical	3	8	49	21	52	19
Modeling	7	19	34	15	41	15
Stat. and Biom. Methods	14	39	62	27	76	28
Total (% within rounding error)	36	100	233	100	269	100

Table 3

Chi-square contribution for each cell of Table 1

Classification	1950-51	1960-61	1970-71	1980-81	Total
Agriculture	0.50	0.67	2.49	0.49	4.15
Animal Populations	0.39	0.03	0.04	0.66	1.12
Assays	3.94	0.00	3.56	0.48	7.98
Biological	0.16	0.42	0.61	2.91	4.10
Clinical Trials	1.58	2.36	13.91	19.47	27.32
Genetics and Breeding	0.52	0.39	0.22	0.30	1.43
Medical	0.97	11.49	3.86	33.78	50.10
Modeling	2.18	2.91	0.02	0.33	5.44
Stat. and Biom. Methods	1.66	1.80	6.03	23.05	32.54

χ^2 (24 degrees of freedom) = 134

$\chi^2_{.01}$(24 degrees of freedom) = 43

Table 4

Contributions to chi-square for row totals of Tables 1 and 2

Classification	Biometrics	Toulouse	Total
Agriculture	4.95	8.59	13.54
Animal Populations	2.73	4.75	7.48
Assays	1.51	2.62	4.13
Biological	0.20	0.34	0.54
Clinical Trials	1.49	2.58	4.07
Genetics and Breeding	2.44	4.24	6.68
Medical	0.02	0.03	0.05
Modeling	2.93	5.09	8.02
Stat. and Biom. Methods	0.21	0.37	0.58

χ^2 (8 degrees of freedom) = 45

$\chi^2_{.01}$(8 degrees of freedom) = 20