SELECTED TOPICS IN BIOMETRY

J. T. Wood
The Australian National University, Canberra, Australia

Keywords: Statistical analysis, designed experiments, spatial analysis, multivariate analysis

Contents

Summary

Variability is an unavoidable element of most biological processes. Sometimes it is of interest in itself; sometimes we need to disentangle it from effects of interest.

A diverse and rich theory has been developed to handle the many types of variability which occur. Sometimes our aim is to summarize a set of data; other times we want to draw inferences from data. These may include testing hypotheses about the how the data were generated and estimating parameters of interest. This article discusses three areas, designed experiments, spatial processes and multivariate analysis, in which distinct bodies of theory have evolved. Some other areas, including variation in time and genetic variation, are discussed more briefly.

1. Introduction

Variability is an intrinsic element in most biological systems and processes. For example, the weights of mice caught in traps will vary from individual to individual; trees in a woodland will not be located in a completely systematic way. Data will often be subject to measurement error, which may sometimes be significant relative to other sources of variability.

The role of biometry is to describe the essential characteristics of this variability, or to enable us to disentangle effects of interest from the variability. We may simply wish to have a convenient summary of a large set of data; or we may wish to establish whether the sizes of observed differences between treatments can be explained as a result of random variation or whether they are better explained as due to systematic differences between treatments. More usefully we may wish to estimate the underlying systematic differences and to establish how accurate the estimates are. In these respects the practice of biometry does not differ greatly from the application of statistics in other areas, although biometry sometimes also refers to deterministic mathematical modeling of biological systems.

Over time a distinction has developed between the application of statistics in medicine and epidemiology, where it is referred to as biostatistics, and other applications of biometry. This article focuses on biometry as applied in these other areas, such as ecology, agriculture, horticulture, forestry, etc.

Biometry originally developed as the summarization and interpretation of large bodies of biological data. However in the first half of the last century Sir Ronald Fisher working at the Rothamsted Experimental Station in England recognized that methods applicable to small data sets needed to be developed for the analysis of much of the data which he encountered there. His work revolutionized the practice of biometry.

Many of the problems in the application of biometrical concepts relate to the complexity of the random structure underlying the data. This can be induced by the way in which the data was collected, or by the intrinsic properties of the process being investigated. Examples of the first include the way in which a field experiment has been laid out or the structure of a sampling scheme used. Examples of the latter include measurements made on a random processes operating over space or time, and multivariate data where several different measurements are made on the same individual. Methods for handling complexity of random structure form the underlying theme of this article. Failure to take proper account of this structure invalidates statistical inferences drawn from the data, and can mean that important effects are overlooked; that negligible or non-existent effects are judged to be important; or that the way in which a process operates is misunderstood.

Developments in computing continue to change our perceptions of what is meant by “large” and “small” datasets, and to increase the complexity of the models for random structure which can be handled. However, from the point of view of making statistical inferences, “small” relates to the number of units for which data is available, rather than the actual volume of data.

2. Inference

Although biometry often involves data sets with quite complex structures, the analysis of the data usually has some or all of only three elements, choosing a good model, testing hypotheses, and estimating various quantities. To be useful our estimates will also include some measure of their accuracy.

2.1. Hypothesis testing

Statistical hypothesis testing addresses the issue of whether or not the observed data are consistent with some assumption about the way they were generated. It can never prove an hypothesis, only lead us either to reject an hypothesis, or to conclude that the data do not give us any good reason to reject it. In the latter case we might conclude that for practical purposes the data are consistent with the hypothesis or else that the data are insufficient to say anything very useful about it.

For example, suppose we have two groups each of six mice, and the weights in grams of the mice in one group are (62.2, 61.9, 54.7, 49.3, 45.5, 58.2) and in the other are (34.5, 49.8, 48.9, 49.5, 46.0, 48.5). The mice in the first group seem to be bigger on average than those in the second, but there is some overlap, so we might wish to test whether the two groups can be considered to be drawn from the same source. The first step is to choose a test statistic which will summarize the observed differences between the groups. A natural choice is the difference between the average weights of the mice in the two groups, although there are many possible alternatives. We have no *a priori* reason to suppose that the means are identical, and we want to see if the observed difference is consistent with the sort of random variation one might expect if the two groups were obtained in the same way. The means for the two groups are 55.3 and 46.2, and the difference is 9.1. A conceptually simple test is a randomization test where we take the twelve weights and consider all possible ways of dividing them into two groups of six. There are 924 ways in all. We can then see how many of these divisions lead to a difference between the first and second groups, which is greater than or equal to 9.1. In this case there are 17 such divisions, which is 1.84% of the total. However, interchanging the two groups, there will also be 17 divisions in which the difference between the second and first groups is greater than or equal to 9.1. So if we were to repeatedly select two groups of six mice at random from the twelve mice we would get a difference as extreme as that observed 3.7% of the time. We have to decide whether 3.7% is sufficiently small as to make us doubt that the data have arisen from some mechanism equivalent to this. We can quote 3.7% as summarizing the evidence against this null hypothesis of no difference, or we can choose some essentially arbitrary number, 5% is a popular selection, and reject the null hypothesis if the proportion of randomizations giving a more extreme value than that observed is less than this. In this case we say that the null hypothesis is rejected at the 5% level.

In all this the null hypothesis only assumes that the mice are all drawn from the same source. The testing procedure does not require any extraneous assumptions. It could equally well be applied if some other statistic than the difference between the two means was chosen as the basis for the test. In practice we would choose a statistic sensitive to the sort of departures from the null hypothesis we are interested in. For

example, if we expect our data to include occasional very large or small values we might choose the difference in the medians (the middle values) of the two groups as our test statistic to minimize the effect of these extreme values. Note that we have assumed that there is no structure in the groups of mice which we could take account of. For example, if we knew that three of the mice in one group were drawn from the same litter, we ought to take account of this in our analysis.

Of course in many situations complete enumeration of the possible permutations of the data is impractical, but we can use a random sample of permutations instead. Often there is a theoretical approximation which involves quite mild assumptions. In many other cases the null hypothesis is more complex, so that a convenient randomization test may not be available. However the same principle of establishing the proportion of values for the test statistic more extreme than the observed value which would be obtained under the null hypothesis applies.

The power of a statistical hypothesis test is an important concept. For a specific alternative to the null hypothesis, it is the probability that the null hypothesis will be rejected when the alternative hypothesis is true. Clearly we would like the power to be as large as possible. There are usually many alternative null hypotheses each with its own value for the power of the test. If we are interested in specific alternative hypotheses, comparing the power for different tests helps us to choose the most appropriate test for our purpose, and tells us whether the data are adequate for detecting the sorts of effects we are interested in. Power considerations are important in deciding how much data to collect, since power will increase as the sample size increases. In the example of comparing two populations of mice we might well decide how many to measure by deciding on the power we would like to have for a specific difference in the average weights in the two populations.

2.2. Confidence Intervals

In the example of the weights of two groups of mice in the previous section we may be more interested in estimating the difference between the average weights for the populations from which the groups are drawn. We can estimate this by the difference in the average weights for the two groups, but this is not usually of much use without some idea of how accurate this estimate is likely to be.

For any given value we can test the hypothesis that it is equal to the difference simply by subtracting it from the values in the group hypothesized to have a larger mean, and then testing for equality at a particular significance level, as described in the previous section. If the hypothesis is not rejected we can say that the value for the difference being considered is consistent with the data. If we take all values which are judged consistent with the data, these values constitute a confidence region for the difference. In most situations such regions constitute a single interval whose upper and lower values define a confidence interval.

In this way we have a duality between confidence intervals and hypothesis testing, since testing whether a parameter takes a specific value or not is equivalent to asking whether that value lies within a confidence region. The percentage level of the confidence

interval is said to be 100 minus the chosen significance level, so if the chosen significance level is 5% we say that we have a 95% confidence interval.

2.3. Model Selection

In the physical sciences theoretical considerations often dictate the choice of a mathematical model or equation to describe the way in which a particular data set has arisen. In biometry this is usually not the case. Consider modeling the growth of an animal which has been measured at a sequence of points in time. We might suppose that initially it will grow rapidly and eventually its size will plateau at some value, and we may be able to suggest some intuitively plausible equation to describe the growth. However the equation will not relate directly to the mechanisms controlling growth. Our choice of possible models will be governed by our purpose in collecting and analyzing the data. Is it simply to interpolate at times for which we do not have measurements? Is it to give us a method for comparing the growth of different individuals, perhaps in different environments? Is it to provide some reference standard for assessing whether the growth of other individuals in some future circumstances has been affected in some way? Is it to see whether the growth consists of distinct phases with different characteristics? Having selected a set of candidate models, our task is to see which, if any, give an adequate description of the data, and, if we have successfully found a good model, to estimate constants or parameters in the equation.

We might think that it is best to choose the model which fits the data most closely. However this may lead to choosing a model which reflects chance peculiarities of our data, and does not reflect the mechanism by which the data arose. In addition, other things being equal, a complex model will fit our data better than a simple model, although the simple model may be much easier to interpret and be a more convenient tool for comparisons with other sets of data.

For this reason statistics, such as the Akaike information criterion, have been proposed which are a composite of the complexity of a model and the degree to which it fails to fit the data, and these can be used to discriminate between different models. Complexity in this context refers to the number of unknown parameters which have to be estimated to fit the model to a set of data. Of course we often have some subject matter knowledge which tells which models make the most sense, and this must also be taken into account, so that our model will reflect a compromise between our prior knowledge, and the results of a formal model selection process.

-
-
-

Bibliography

Cressie, N.A.C. (1993) *Statistics for spatial data*, New York: Wiley [Comprehensive account of methodology for analysis of spatial data as of 1993]

Digby, P.G.N. and Kempton, R.A. (1987) *Multivariate analysis of ecological communities*, London: Chapman and Hall [A good account of some of the multivariate methods used in ecology, with sound advice about the use of classification and ordination techniques]

Diggle, P.J. (1983) *Statistical analysis of spatial point patterns*, London: Academic Press [An approachable introduction to this topic]

Diggle, P.J. Heagerty, P.J. Liang, K-Y. and Zeger, S.L. (2002) *Analysis of longitudinal data*, 2nd edn, New York: Oxford University Press [An important recent book emphasizing repeated measures ideas]

Falconer, D.S. and Mackay, T.F.C. (1996) *Introduction to quantitative genetics*, 4th edn, London: Longman. [A classic text on this topic]

Fisher, R.A. (1966) *The design of experiments*, 8th Edition, Edinburgh: Oliver and Boyd [The seminal reference for design of experiments]

Gordon, A.D. (1999) *Classification*, 2nd Edition, Boca Raton: Chapman and Hall/CRC [A good recent account of the current state of classification methodology]

John, J.A. and Quenouille, M.H. (1977) *Experiments: design and analysis*, London: Griffin [An excellent development of the concepts underlying designed experiments]

John, J.A. and Williams, E.R. (1995) *Cyclic and computer generated designs*, 2nd Edition, London: Chapman and Hall [A good recent account of experimental design. The content is broader in scope than the title would suggest]

Krzanowski,W.J. (1988) *Principles of multivariate analysis: a user's perspective*, (reprinted with corrections, 1990) Oxford: Clarendon Press [Comprehensive account of multivariate analysis, but with strong emphasis on strategy for analysing data]

Mardia, K.V. Kent, J.T. and Bibby, J.M. (1979) *Multivariate analysis*, London: Academic Press [Excellent comprehensive account of multivariate analysis]

Webster, R. and Oliver, M. (2000) *Geostatistics for environmental scientists*, Wiley [A good account of methods for random spatial processes at an accessible level]

Biographical Sketch

Jeff Wood has an M.A. in Mathematics from Cambridge University, an M.Sc. in Statistics from the University of Wales and a Ph.D. in Mathematical Statistics from the University of Birmingham, U.K. He joined the Statistics Section at the National Vegetable Research Station at Wellesbourne near Warwick in the U.K. in 1966.

In 1973 he moved to the Division of Mathematical Statistics of the Commonwealth Scientific and Industrial Research Organisation in Canberra, Australia and remained there through various reorganizations and changes of name until 2001, when he moved to the Statistical Consulting Unit of the Australian National University.

Throughout his career he has applied statistical methods to many disciplines including ecology, agronomy, horticulture, forestry, soil science, and road safety research

He has served the Statistical Society of Australia and the International Biometric Society in various roles. He is currently Treasurer of the International Biometric Society. He is an Accredited Statistician of the Statistical Society of Australia.