CERTAIN MULTIVARIATE DISTRIBUTIONS IN THE PRESENCE OF INTRACLASS CORRELATION

By B. M. Bennett

University of Washington, Seattle, U.S.A.

1. Introduction

First we include a brief restatement of certain well-known results on distribution theory in multivariate analysis, in particular those concerned with the Wishart and generalized T^2 distributions.

Suppose $\{y_{i\alpha}\}$ is a sequence of n observations each from a p-variate normal population $(i = 1, \ldots, p;\ \alpha = 1, \cdots, n)$

$$\frac{|\Lambda|^{\frac{1}{2}}}{(2\pi)^{p/2}} e^{-\frac{1}{2}\sum_{i,j} \lambda_{ij} (y_i - \xi_i)(y_j - \xi_j)} dy \tag{1}$$

with means $E(y_{i\alpha}) = \xi_i$ $(i = 1, \ldots, p)$, and unknown variance-covariance matrix $= \underset{\smile}{\Sigma} = \| \sigma_{y_i y_j} \| = \| \sigma_{ij} \| = \| \lambda_{ij} \|^{-1} = \underset{\smile}{\Lambda}^{-1}$.

If the vector consisting of the p sample means $\underset{\smile}{\bar{y}} = (\bar{y}_1, \cdots, \bar{y}_p)$ is defined by $n\bar{y}_i = \sum_\alpha y_{i\alpha}$ $(i = 1, \ldots, p)$ and $\underset{\smile}{A} = \| a_{ij} \|$ denotes the matrix of sums of squares and cross products, $a_{ij} = (n-1) s_{ij} = \sum_\alpha (y_{i\alpha} - \bar{y}_i)(y_{j\alpha} - \bar{y}_j)$ it is known that the vector $\underset{\smile}{\bar{y}}$ also has the multivariate normal distribution (1) with now the matrix $\underset{\smile}{\Sigma}$ replaced by $1/n\, \underset{\smile}{\Sigma}$. Furthermore the sequence $\underset{\smile}{\bar{y}}$ and the a's are independently distributed, and it is known that the a's have the Wishart distribution[3] with $(n-1)$ degrees of freedom,

$$\frac{|\Lambda|^{\frac{1}{2}(n-1)}}{C(n,p)} |A|^{\frac{1}{2}(n-p-2)} e^{-\frac{1}{2}\sum_{i,j} \lambda_{ij} a_{ij}} da \tag{2}$$

the constant of integration being $C(n,p) = 2^{\frac{1}{2}p(n-1)} \pi^{\frac{1}{4}p(p-1)} \prod_{i=1}^{p} \Gamma(n - i/2)$.

As a test of the hypothesis H_0: $\xi_i = \xi_i^0$ $(i = 1, \ldots, p)$ Hotelling proposed the generalized T^2 statistic[1]

$$T^2 = \sum_{i,j=1}^{p}\sum s^{ij} d_i d_j \tag{3}$$

if we set: $d_i = \sqrt{n}\,(\bar{y}_i - \xi_i^0)$, $(i = 1, \cdots, p)$, and demonstrated that $T^2/n - 1$ has the same distribution as the ratio of two independent chi-squares with p and $(n - p)$ degrees of freedom, respectively. In particular, the distribution may be written

$$\frac{1}{B\left(\frac{p}{2}, \frac{n-p}{2}\right)} \frac{\{T^2/n - 1\}^{\frac{1}{2}(p-2)}}{(1 + T^2/n - 1)^{\frac{1}{2}n}}\, d\,(T^2/n - 1). \tag{4}$$

2. Effect of Intraclass Correlation

Suppose we again consider a sequence $\{y_{i\alpha}\}$ of observations on the p characters as defined before, but in addition we assume that:

$$\sigma_{y_{i\alpha}\, y_{i\beta}} = \rho_i \sigma_{ii}\ (\alpha \neq \beta)\ (i = 1, \cdots, p)$$

in designating the intraclass correlation coefficients $(= \rho_i)$. Concerning the sequence $\{y_{i\alpha}\}$, we assume a model of the form:

$$y_{i\alpha} = u_i + v_{i\alpha} \tag{5}$$

if $\{u_i\}$ $(i = 1, \cdots, p)$ denotes a sequence of p independent normal variates with means $E\,(u_i) = \xi_i$ $(i = 1, \cdots, p)$ and a common variance $= \sigma_u^2$, and $\{v_{i\alpha}\}$ a sequence from a multivariate normal distribution with means equal to zero and a certain unknown variance-covariance matrix. We also assume that the u and v sequences are independent.

In particular, we note from (5) that

$$\begin{aligned} \sigma^2_{v_{i\alpha}} &= \sigma^2_{y_{i\alpha}} - \sigma^2_{u_i} \\ &= \sigma_{ii} - \sigma_{y_{i\alpha}\, y_{i\beta}} = (1 - \rho_i)\, \sigma_{ii} \\ \sigma_{v_{i\alpha} v_{j\alpha}} &= \sigma_{ij} \quad (i \neq j) \end{aligned} \tag{6}$$

As before we denote $a_{ij} = (n - 1)\, s_{ij} = \sum_\alpha (y_{i\alpha} - \bar{y}_i)\,(y_{j\alpha} - \bar{y}_j)$ $= \sum_\alpha (v_{i\alpha} - \bar{v}_i)\,(v_{j\alpha} - \bar{v}_j)$ and using the definitions (6) and that of the a's we have the result

Theorem 1.—*If* $\underset{\smile}{A}^* = \|a_{ij}^*\|$, *where* $a_{ii}^* = a_{ii}/(1 - \rho_i)$ $a_{ij}^* = a_{ij}$ $(i \neq j)$, *then the* a^**'s have the Wishart distribution* (2) *with* $(n - 1)$ *degrees of freedom.*

Concerning the sequence $\underset{\smile}{\bar{y}} = (\bar{y}_1, \cdots, \bar{y}_p)$ of the sample means, it is easy to verify from the properties of orthogonal transformations that $\underset{\smile}{\bar{y}}$ has the multivariate normal distribution (1), except that the matrix $\underset{\smile}{\Sigma} = \|\sigma_{ij}\|$ is replaced by $1/n\, \underset{\smile}{\Sigma}^* = 1/n\, \|\sigma_{ij}^*\|$, where σ_{ii}^*

$= \{1 + (n - 1) \rho_i\} \sigma_{ii}, \sigma_{ij}^* = \sigma_{ij} (i \neq j)$. Furthermore $\bar{y}$ and the a^*'s are independently distributed.

The appropriate modification of the T^2 test may be made as follows. Define $d_i^* = \sqrt{n} (\bar{y}_i - \xi_i^0)/\sqrt{1 + (n - 1) \rho_i}$, $s_{ij}^* = a_{ij}^*/n - 1$ $(i, j = 1, \cdots, p)$. Then it may be stated that

Theorem 2.—*The statistic* $T^{*2} = \sum_{i, j=1}^{p}\sum s^{*ij} d_i^* d_j^*$ *also has the distribution* (4).

A confidence set with confidence coefficient $= 1 - \epsilon$ for the unknown vector $\underset{\smile}{\xi} = (\xi_1, \cdots, \xi_p)$ if the ρ's are known also may be obtained from the region: $(n - p) T^{*2}/p (n - 1) \leqslant F_\epsilon$, if F_ϵ represents the critical value of the F or variance ratio test with a level of significance $= \epsilon$ and with p and $n - p$ degrees of freedom respectively.

Results similar to Theorems 1 and 2 may be obtained for the case of two multivariate samples and for the more general regression model similar to that expressed by (5).

3. Summary

In the multivariate distribution theory developed thus far it has been customarily assumed that the successive observations on each of the p multivariate characters are independently obtained. Occasionally in the course of certain biological or agricultural experimentation where proximate measurements on many variables are involved such an assumption must in reality be relaxed. This paper gives a brief exposition of the effects of 'intraclass correlation' on several of the principal multivariate sampling distributions, as a generalization of Walsh's results.[2]

4. References

1. Hotelling, H. .. "Generalizations of Student's ratio," *Ann. Math. Stat.*, 1931, **2**, 359–78.

2. Walsh, J. E. .. "Concerning the effect of intraclass correlation on certain significance tests," *ibid.*, 1947, **18**, 88-96.

3. Wishart, J. .. "The generalized product moment distribution in samples from a normal multivariate population," *Biometrika*, 1928, **20**, 32–52.

INFLUENCE OF AGRONOMIC INVESTIGATION ON THE SCIENCE OF STATISTICS*

BY JOHN WISHART

THERE was a time when mention of the word Statistics conjured up nothing more than a picture of Government offices compiling blue books of factual data. By now, however, it is becoming pretty generally understood that there is a Theory of Statistics, this term defining a subject which has an established place among the range of scientific subjects studied at most universities. In particular, the subject of mathematics has kept abreast of modern developments, and has gained in power and prestige, by developing a branch dealing with Mathematical Statistics, which is studied both as Pure and as Applied Mathematics. Now in no direction has the Statistical Method been applied more, during the last 30 or more years, than in that of Agricultural Science, a science which advances by dint of much hard work directed towards the conducting of experiments in the field. These years have seen improvements and refinements in the experimental technique of the agronomist, and it has been common for statisticians, some if not all, to claim that their Science has had a very marked influence on the development of experimental agriculture. It is, of course, true to say that mathematicians well grounded in theoretical statistics have been imported into Agricultural Colleges and Research Stations, and that, with their help, agronomic investigators have learnt, sometimes slowly and usually painfully, to carry out their experiments in a particular way, and to perform series of arithmetical computations which seem to become more and more complex as time goes on. But tonight I would like to take you back with me a little way into the history of this development, and to advance a different thesis. This will be to the effect that agronomic investigation has had a most profound effect on the building up of the Science of Statistics as we know it today. The agronomists have helped to create a subject with a most valuable discipline of its own for educational, *i.e.*, mind-training purposes, and to open a wider field for research and further application. Others should join the statisticians in this tribute, for the methods developed have proved to be of signal benefit in many directions, not least in the study of industrial processes and productivity.

---

* Lecture given at the Eighth Annual Meeting of the Indian Society of Agricultural Statistics, November 28, 1954.

Some one hundred years ago, Lawes and Gilbert were experimenting at Rothamsted with chemical fertilizers as a substitute for organic manures. The results were promising, but there were some who thought that in the course of time the lack of organic matter in the soil, or the acumulation of inorganic chemical substances, would have a deleterious effect on yield. To prove them wrong, Lawes and Gilbert continued their experiments annually, and these experiments are still in being today in substantially their original form. When a statistical department was founded at Rothamsted Experimental Station about 1919, a most valuable collection of experimental records was available. Among other things, their study led R. A. Fisher to useful methods for the analysis of series of annual data, including the separation of slow time changes from periodic fluctuations and random residual effects. The methods of orthogonal polynomial fitting, as we know them today, were vastly improved both in their theoretical formulation and computational technique, because the data were there to be analysed. Contributions to the theory of correlation resulted from the same studies. In particular, the random sampling distribution of the coefficient of multiple correlation, in the special case of the population coefficient being zero, was reached as a byproduct of the study of the influence of rainfall on the yield of wheat at Rothamsted, and this undoubtedly stimulated Fisher to produce later the sampling distribution in the general case. Periodogram analysis came in for its share of attention because of this same data.

Three features stood out in the data available for examination at this time. One was that, even for experiments which had continued over a number of years, the available samples of data were small, so that it was doubtful whether the existing large sample battery of tests, produced over many years by the labours of Karl Pearson, was adequate for the purpose. A second was that the experimental errors met with in field experimentation were very large by comparison with those customary in other scientific fields, for example astronomy. A third related to the fact that a heavy contributor to the error was the variable soil fertility, and that this was not random in incidence, but often had semi-permanent effects. So that even although at Rothamsted there were many years' records of yields of plots, laid out side by side with different manurial treatments, it seemed impossible to say how much of the yield difference was due to the difference in treatment, and how much to the soil differences. It will never be known whether the small sample theoretical studies that have so much advanced the science of statistics would have been made anyhow had the mathematician not been challenged by the agronomist to produce solutions. In fact, however, the difficulties were there, the challenge was made, the right men were available to take it up, and the results are today in the text-books.

Let me amplify this a little. "Student" was actively engaged with the data of barley variety trials in Ireland when he saw that he could never hope to have large samples, and that he should therefore take account of the uncertain estimates made of the error in testing for the significance of differences between variety means. His epoch-making paper of 1908 produced what is now known as the *t*-test, or Student's test, which has become a household word in statistics. Not only was this contribution valuable in itself, but the idea of what has since become known as studentization led other workers to examine and develop other tests, with the result that many exact sample tests took the place of earlier approximate ones. This possibly did more than any other single development to give the theory of statistics a recognition in the domain of mathematics, because it could now quite genuinely be called mathematics, a thing about which many were doubtful before.

Around the years 1910 and 1911 it was the experimental activities of the agronomists that produced valuable records of uniformity trial data. These were distinguished from the earlier Rothamsted records because the time, or year-to-year, effect did not enter, and because all plots were uniformly treated, so that it was possible to study both systematic and random variations in yield. Systematic soil fertility changes were recognised and plotted, and likewise tabulations were made of the standard errors likely to be met within plots of varying sizes. It was seen that the smallest size plots had too large an error, but that on the other hand if the plot size was made large, then large systematic variations were to be expected across the area of any projected experiment. A compromise was arrived at with an intermediate size of plot, and the next thing was to devise suitable experimental arrangements. The work done at that time was often heavily criticized later because the designs adopted were of a systematic nature, but perhaps it was not the technique which was weak so much as the apparatus available at the time for analysing the results. I say this because within recent years systematic designs for experimentation have been seriously advocated, provided special methods are adopted. However this may be, the methods used up to 1925 did recognise that the different treatments imposed on the crops should at least roughly sample areas of equal soil fertility and investigators began to use what is now called replication, meaning by this that a set of plots having variable treatments should be repeated, and perhaps repeated again, on the area of ground given up to the experiment. One very popular form of experiment at that time, for two varieties of grain crops, was Bevan's Half-drill Strip Method. The very simple device of blocking the middle of a seed drill, and filling it to left and right with different varieties, *A* and *B*, say, led to the automatic plot arrangement *ABBAABBA*.... as the drill went up and down the field. This method led to another theoretical statistical advance. The two varieties being compared were close together in narrow plots, and the difference $B - A$ could therefore be assumed to be relatively uncontaminated with fertility differences. Even if there were some trend of fertility, the fact that the order was reversed in the next pair of plots produced an automatic correction. Furthermore, there might be 10 or more such differences in the average experiment, so that the mean difference would have the lower standard error appropriate to the mean of a fairly substantial number of pairs of yields. Now, not only was this good experimental practice, except for the one possible point of criticism in that the arrangement was systematic throughout, but also it led to the new well-known modification of Student's test known as the test for paired comparisons. The method, in fact, originated the idea of a block of plots, in this case consisting of two neighbouring strips, which was to be repeated a number of times. Bevan, the inventor of the arrangement, was an experimental agronomist, and I believe that the idea was his own, although "Student" came in with the necessary modification of his own method of analysis. My point is that I believe it to be right to say that a statistical method which has since been found to be most useful in other fields, and which proved to be the germ from which came the more elaborate methods of analysis of variance, was largely brought into being by the experimental investigations of the agronomist.

When more than two varieties or treatments were incorporated in a single experiment, the idea of the block persisted, for it seemed natural to set the treatments out in order *ABCD*.... ; and then repeat, possibly in the reverse order to correct for fertility trends. A variation was to insert a control, which in a variety trial might be a standard variety about which a good deal was known, or in a manurial trial might be a plot with no treatment at all, between two treated plots, or alternatively every few plots along the row. The need to analyse the yield data from such experiments stimulated the statistician, not only to study the best experimental arrangements which would keep the error down while at the same time permitting its valid estimation, but also to examine what the right generalization was to the method of paired comparisons. There was a practical point here in that, with, say, five treatments, there were 10 possible ways of comparing one treatment with another, and not only would this be a tedious operation by "Student's" methods, but also it would not be very accurate because of the limited number of repetitions of any one pair. What I am leading up to is, of course, that the study of these problems led Fisher to evolve the statistical method generally known as the analysis of variance, together with a varied battery of tests based on his discovery of the random sampling distribution of the ratio of two independent estimates of variance. It is true that Fisher's work fundamentally affected agronomic experimentation for the future because he himself suggested the methods of Randomized Blocks, Latin Square and the like, but it is at least open to speculation whether this important section of the theory of statistics would yet have evolved but for the influence of the agronomists and their problems.

The theory of estimation and of tests of significance may therefore be said to have received great stimulus from the practical work and needs of the agronomist. Given this start, *i.e.*, with a problem to solve, it was natural to expect that gifted mathematicians would go on to examine the foundations of the methods so far used, and the assumptions made, and so build up a sound mathematical structure for statistical theory which would be no mean part of a course on mathematical analysis. More than one branch of mathematics have seemed to become sterile, until some person or group came along with a problem of a new character to solve, whereupon this led to a fresh outburst of activity on the part of at least one group of mathematicians. So it has been in the present generation with mathematical statistics, until today we find that in every monthly issue of *Mathematical Reviews* a substantial part is taken up by the sections Probability and Mathematical Statistics. In addition, though not so frequently, there are sections on Mathematical Biology and Mathematical Economics. It will be found also that most teachers of mathematical statistics refer over and over again for illustrative purposes to the experimental activities of the agronomist.

It was not long before the experimentalists felt the need to include more than one factor in the same experiment. By this I mean that they might wish to combine certain manurial tests with a variety trial, since a particular variety might prove only to be the best under a particular manurial treatment. Or they might be seeking for the best combination of quantities of the usual applied nutrients containing nitrogen, potash and phosphate. A set of varieties, or a set of applications of one fertilizer, is what I mean by a factor. This need was met by designing factorial experiments in which each level of one factor was in combination with all levels of the others. Again mathematical problems were met with here, and the study widened to include that of mathematical models as underlying assumptions, and finally to a whole range of problems and tests comprised in the general title of Design of Experiments, courses in which are given in most mathematical schools where statistics is taught. It has been argued that in this field the mathematician has outrun the needs of the experimenter, but that was because the subject was a fascinating one in itself to a pure mathematician, who wanted to pursue it for its own sake. We can, I think, say that this is as much mathematics as some other branches of the subject that could be mentioned, and it owes its origin to the investigations of the agronomist. This aspect of the subject has been found to interest the more abstruse type of mathematician who might hitherto have studied the theory of groups, or combinatory analysis, or even the theory of numbers. The consequent interaction that then takes place between mathematicians of different fields of primary interest is undoubtedly good for mathematics. On the applied side, the increase in the number of factors tested in agronomic experiments, together with the development of confounding, a name given to a special kind of lay-out in which a complete replication is divided into block groups to eliminate some of the fertility variations while at the same time permitting the valid assessment of most of the important effects, has provided just what was required in the industrial experimentation sphere, since in that work there seems no end to the possible variations that have to be studied in industrial processes among the numerous factors involved.

The further quest of the agronomist for greater accuracy in his experiments, particularly those conducted with farm animals, stimulated the mathematician to see whether anything could be done along the lines of regression or correlation analysis. The small sample tests of significance of "Student" had some time before been extended to evolve similar tests for the regression coefficients, simple, partial and multiple, and the analysis of variance had been exploited to cast such tests into an easily comprehended form. The position then was that there was a relatively finished structure for the analysis of multiple classification data in one variable, but that in regression analysis, *i.e.*, where more than one variable entered in, the methods were confined to homogeneous data. The next step was the logical one of studying regression analysis for data in multiple classifications in two and more variables. Thus, in response to practical needs, a method of combined analysis of variance and covariance was evolved, which again provided an interesting and provocative stimulus to the mathematician. The logical end to this was, of course, the development of general methods of multivariate analysis, which can be regarded as the highwater mark today in the flow of mathematical-statistical ideas.

All these developments are inseparable from the name of one man, R. A. Fisher, and it could be argued that it is impossible to keep a good man down, and that if his career had been in different sphere his own work, and that of the pupils he inspired, might have been equally fruitful in the contributions they made to mathematics, if not to statistics. But we are talking about the Science of Statistics, and I can think of only one possible field other than agronomy, namely that of physical anthropology, which during this period was furnishing its data and throwing up its problems for mathematical solution. I exclude economics deliberately, since it is only in very recent years that work in that field has directly added to the apparatus of mathematical statistics. The field of physical anthropology was worked on by Karl Pearson in England and by Mahalanobis in India, and influenced the trend of statistical science in certain specialised directions. But the total effect, while important, especially in the domain of multivariate analysis, does not match up to the stimulus provided by the agronomic investigator. I trust, therefore, that it will be conceded that I have made good my thesis, namely, that the experimental activities of the agricultural research worker have, during the present century, exerted a most profound influence on a growing science, namely that of Statistics.

It is fitting that I should end by paying a tribute, on behalf of numerous mathematical statisticians, to the agronomists for the stimulus and inspiration provided by their labours. I would couple this with an appeal to them to continue the good work, were it not for the fact that the proceedings during the present session have provided ample evidence that no such appeal is necessary. I think we, as receivers of favours, should give something back, and, if I may, I would like to make a suggestion which I hope will be well received by everyone, and which arises from my observations of the work done in experimental design in India, as well as in other countries. India has perhaps more than its proportionate due share of gifted mathematicians, many of whom have contributed materially to the study of experimental design. It is natural that they should be interested, for its own sake, in the *theory* of design, and in generalizations rather than in particular cases. What I am asking is that, in giving instruction to those whose duties will lie in the agronomic sphere, they will realise that they are not talking to mathematicians, but will be patient in explaining first principles, will concentrate on the simpler forms of design that are of very general application, and will take time to expound the practical aspects of the necessary statistical analysis of the resulting data, which is a very necessary follow-up to the work of choosing a design.