Proceedings of the 1984
Winter Simulation Conference
S. Sheppard, U. Pooch, D. Pegden (eds.)

MULTIVARIATE SIMULATION OUTPUT ANALYSIS: PAST, PRESENT, AND FUTURE*

Linda Weiser Friedman
Assistant Professor
Department of Statistics and Computer Information Systems
Baruch College of the City University of New York
New York, N.Y. 10010

ABSTRACT

The overwhelming majority of research into the design and analysis of simulation assumes that only a single response variable is of interest to the researcher. This is an impractical assumption. In these introductory remarks, the problem of multiple response in simulation analysis is addressed and some solutions are discussed.

INTRODUCTION

For some time now, researchers have investigated the design and analysis of uni-response simular experiments. However, it is a rare system simulation which outputs only a single measure of effectiveness for analysis. The researcher conducting a simulation experiment will probably find that results must be analyzed on more than one dimension. These measurements will, in all likelihood, be correlated. Thus far, research into the statistical design and analysis of multiple response simular experiments has been sparse.

There is a fair-sized body of scientific papers published since the late 1960's which is concerned with the statistical analysis of simulation output data (7). The overwhelming majority of this body of work ignores the multivariate aspect of simular output data and approaches the various simulation concerns from the perspective of a single output measure. Until very recently, the few researchers who did take notice of the problem of multiple response have, for the most part, mentioned it and then avoided it. That is, the assumption was made, either implicitly or explicitly, that there was only one response variable of interest to the researcher. Even some otherwise complete textbooks such as Gordon (8) and Fishman (6) ignore the multiple response problem.

SOME SOLUTIONS

A small handful of authors have attempted to analyze multiple response simular data using techniques which, as we shall see, are inferior to the multivariate statistical methods. Two techniques which have been proposed to solve the multiple response problem are: performing several univariate analyses on the same set of data and combining the responses into a single response function. These two were presented first by Naylor et al (24, 25), then by Hunter and Naylor (12), and subsequently by Shannon (31).

A general problem with performing several univariate analyses on the same set of data is that it does not take into account the interdependence among the response variables. This interdependence is almost always present in system simulation experiments. When several univariate hypothesis tests are performed on the same set of data and no adjustment is made to the significance level, there is an additional problem. The problem is that when several tests are performed, each at a particular significance level, say $\alpha=.10$, the significance level for the whole study (i.e., the experimentwise error rate) does not necessarily remain at an alpha of .10. The probability of at lease one of these univariate tests producing a significant result, when indeed only random variation is present, increases greatly as the number of individual tests increases.

The alpha (Type I) error is the probability of rejecting the null hypothesis when it is true. If only one statistical test is performed, an alpha level of .10 results in a 90% confidence coefficient ($(1-\alpha)$x100%) for the study. If, however, several univariate tests are performed on the same set of data, the actual experimentwise error rate ranges from a low of alpha, if the different measures are perfectly correlated, to a maximum value of $1-(1-\alpha)^p$, if the measures are mutually independent, where p is equal to the number of measures for which individual univariate statistical tests are performed. Suppose four (p=4) simultaneous statistical tests are performed, each at a significance level of $\alpha=.10$, on four responses measured from the same data set (a simulation experiment). The experimentwise alpha level may be as high as $1-(.9)^4= .34$, hardly an appropriate significance level by anyone's standards.

One old and commonly used approach is to scale down the univariate significance level in order to maintain the desired experimentwise error rate. This may be done either via the so-called Bonferroni approach (see 16, 21) in which the individual significance levels are set to α_E/p, where α_E represents the desired experimentwise error rate, or by algebraically manipulating the formula

---

*taken in part from the author's doctoral dissertation submitted to the Faculty of the Polytechnic Institute of New York in partial fulfillment of the requirements for the degree Doctor of Philosophy (Operations Research), 1983.

$\alpha_E = 1-(1-\alpha)^p$ in order to solve for α. Thus, with p=4, if the experimentwise error rate is to be set at .10, the univariate alpha level would be set at α=.025.

It should be noted at this point that this is an overly conservative approach. The chance that one will commit a beta (Type II) error -- that of falsely accepting the null hypothesis -- is increased, and the power of the test is reduced. After all, since the worst is assumed, that is, mutual independence (which is not usually the case), one is obliged to employ an extremely low univariate alpha level. In the example cited above, it will usually be difficult to reject the null hypothesis when testing at a significance level of .025. For the not uncommon situation where p=10 and the desired experimentwise error rate is .05, the individual alpha must be set at .005, an even more unreasonable significance level.

There is another problem with performing several univariate tests on the same set of data. Sometimes, one may find that each of the probabilities (that H_0 is true) associated with the ten different univariate tests falls short of the significance level required for rejection. Thus one might erroneously conclude that all of the null hypotheses were true. However, were a multivariate statistical test performed on the data, one might find that the multivariate null hypothesis should indeed be rejected. How can this happen? To draw an analogy, while getting two of a kind in poker may not be unlikely, getting it ten times in a row may be extremely improbable.

Sometimes, the multiple response problem is "solved" by combining the various responses in some fashion (e.g., linearly) into a single outcome variable by means of a "criterion" or "utility" function, using subjectively-assigned weights. A general method for assigning these weights is given by Kotler (17). This, of course, eliminates the multiple response problem entirely. However, as with any aggregation procedure, much valuable information may be lost. In addition, when weights are assigned subjectively to the responses of a simulation experiment, the researcher must provide strong justification for his choice of weights and extreme caution must be exercised in the analysis and interpretation of any results of the study. Montgomery and Bettencourt's (22) extension of response surface methodology to handle multiple responses made use of just such a utility function.

MULTIVARIATE STATISTICAL METHODS

There has been rare mention in the simulation literature of multivariate statistical analysis. Research into the application of multivariate statistical techniques to the analysis of multiple response simular experiments is virtually nonexistent. A few texts, such as Mihram (20,p.396), Kleijnen (15,p.407), and Shannon (32,p.227), have raised the possibility that these techniques may, indeed, be applicable to simulation.

McArdle (18) presented a case for the application of log-linear modeling, a discrete multivariate method, to the analysis of Monte Carlo simulation data (e.g. robustness studies). The first simulationists to attempt the use of the multivariate analogue of the two-sample t-test were Balci and Sargent (2) and Schruben (29). Seila (30) developed a multivariate estimate of the mean response for regenerative simulations.

It has been argued (15, p.407) that just as the classical univariate statistical analysis techniques of physical experimentation have been found suitable for use in the analysis of uni-response simular experiments, so will the tools of multivariate statistical analysis prove to be suitable for the analysis of multiple response simular experiments.

Consider, for example, the commonly used simular experiment of comparison. This type of experiment may be designed to test the effects of various treatments (e.g., different policies) on the responses of the simulated system or to compare two or more alternative systems with regard to certain measures of effectiveness. When more than one measure of effectiveness is used, a multivariate statistical test should be considered. Treating a simulation with p response variables as p experiments with one response variable each is inferior to multivariate analysis which not only improves the significance level but also enables the researcher to get a better idea of how the responses behave together. Some multivariate techniques which are appropriate for experiments of comparison include the two-sample Hotelling's T^2 test and multivariate analysis of variance (MANOVA).

RATIONALE

Any argument in favor of the applicability of multivariate statistical methods to simulation analysis must be based upon four major points: the experimental nature of simulation; the multivariate central limit theorem; the robustness of certain multivariate statistical techniques; and alternatives, when assumptions of sensitive tests are violated. These four points serve to demonstrate that multivariate statistical methods may, indeed, be applied to simulation output data, with the same degree of caution one would exercise in univariate statistical analysis.

The Experimental Nature of Simulation

There is no doubt that computer simulation is an experiment. The only difference between simulation and physical experimentation is that in simulation, the experiment is conducted with a model of the real system rather than with the real system itself.

Although the simulationist does not have access to the real-world system, simulation is in many ways a superior experimental technique. In simulation, exact replication of experimental conditions is possible by recording and reusing the same random number streams. In simulation, the experiment may be stopped and restarted at the will of the experimenter with no loss of data. Also, since the computer is a frequently used tool for simulation, the real time frame of the experiment may be speeded up considerably, and the risk of (human) measurement error is reduced.

In simulation, whether uni-response or multi-response, the independence of successive replications is maintained by using a different, randomly selected seed value for each replication of the simulation.

The Multivariate Central Limit Theorem

Many univariate statistical analyses of simulation output data rely heavily on the central limit theorem as a means of satisfying the assumption of normality. Similarly, multivariate analyses of simular data will rely on the multivariate central limit theorem.

Multivariate extensions of the central limit theorem, found in Ito(13) and Puri and Sen (28, p.24-25), state that vectors of sample means follow a multivariate normal distribution if the sample size is sufficiently large.

Robustness of Multivariate Statistical Techniques

The robustness of a statistical test is a measure of how well the test stands up under violations of the assumptions with which it was originally developed. Most multivariate statistical tests are derived under the assumption of multivariate normality. Tests of equality of two or more groups, such as Hotelling's T^2 and MANOVA, have the additional assumption that the populations from which the groups were sampled share a common variance-covariance matrix. In truth, there is almost no set of data (simular or otherwise) which meets these assumptions perfectly.

There is a large body of work which demonstrates that the univariate tests upon which many multivariate tests are based are extremely robust under violation of assumptions (except, perhaps, for very small and/or unequal sample sizes and for one-tail tests). Harris(9, p.232) gives an intuitive justification for his conviction that multivariate generalizations of univariate tests will ultimately be shown to exhibit the same robustness to violations of assumptions as do their older, univariate counterparts.

Ito and Schull (14) have shown that, for equal sample sizes, the true significance levels for the two-sample Hotelling's T^2 test match the nominal levels quite well even under violation of the assumption of homogeneity of variance-covariance matrices. Mardia (19) reviewed several robustness studies of the Hotelling's T^2 statistic, and found positive results in all cases.

Other research (5,10,11,3,19) supports the evidence that the T^2 statistics (the two-sample more than the one-sample test) is quite robust against violation of the distributional assumption and the assumption of homoscedasticity.

An extensive robustness study of six MANOVA procedures was done by Olson (26). The procedures were compared with regard to power to detect true differences when various assumptions were violated. Olson concluded that while the "largest root" test should be avoided, others have adequate power to detect true differences.

Alternatives, When Assumptions of Sensitive Tests are Violated

In any particular experiment, if there is doubt as to the applicability of these tools to the data at hand, some very useful tests may be performed to see if the data satisfy the assumptions in question. Andrews et al (1), Mardia (19), and Popoviciu and Vaduva (27) discuss procedures for determining whether the assumption of multivariate normality has been met. There are also tests for the equality of several variance-covariance matrices (9, p.85; 23, p.252).

If the conclusion is that an assumption was not satisfied, then there are two paths open to the experimenter: He may rely on the robustness of the test and use it as is, or he can use an alternate procedure. For extreme departures from multinormality, especially for tests which exhibit sensitivity to extremely non-normal data , Andrews et al (1) discuss the possibility of using data transformations so that the data more nearly satisfy the multivariate normality requirement. While others (4, 19) have attempted to simply revise the offended statistic, more research is still needed in this area.

CONCLUSION

It is hoped that the preceeding will prove to be of some value to the simulation practitioners and researchers who will use multivariate statistical methods in the analysis of simulation output data. There is a great potential benefit awaiting simulationists who use these techniques for the analysis and interpretation of multiple response simulation experiments.

REFERENCES

1. Andrews, D.F., R. Gnanadesikan, and J.J. Warner, "Methods for Assessing Multivariate Normality," In P.R. Krishnaiah (ed.), Multivariate Analysis III, Academic Press, New York, 1973, pp.95-116.

2. Balci, Osman and Robert G. Sargent, "A Methodology for Cost-Risk Analysis in the Statistical Validation of

Simulation Models," Communications of the ACM, vol. 24, no. 11, April 1981, pp.190-197.

3. Chase, G.R. and William G. Bulgren, "A Monte Carlo Investigation of the Robustness of T^2," Journal of the ASA, vol. 66, no. 335, September 1971, pp.499-502.

4. Eaton, M.L. and Brandley Efron, "Hotelling's T^2 Test Under Symmetry Conditions," Journal of the ASA, vol. 65, no. 330, pp.702-711, June 1970.

5. Everitt, Brian S., "A Monte Carlo Investigation of the Robustness of Hotelling's One- and Two-Sample T^2 Tests," Journal of the ASA, vol. 74, no. 365, pp. 48-51, March 1979.

6. Fishman, George S. Principles of Discrete Event Simulation, Wiley, New York, 1978.

7. Friedman, Linda W. and Hershey H. Friedman, "Statistical Considerations in Computer Simulation: The State of the Art," Journal of Statistical Computation and Simulation, vol. 19, no. 3, pp.237-263, 1984.

8. Gordon G. System Simulation, Prentice-Hall, Englewood Cliffs, New Jersey, 1978.

9. Harris, Richard J. A Primer of Multivariate Statistics, Academic Press, New York, 1975.

10. Holloway, Lois N. and Olive J. Dunn, "The Robustness of Hotelling's T^2," Journal of the ASA, vol 62, no. 317, pp.124-136, March 1967.

11. Hopkins, J.W. and P.P.F. Clay, "Some Empirical Distributions of Bivariate T^2 and Homoscedasticity Leprokurtosis," Journal of the ASA, vol. 58, no. 304, pp.1048-1053, December 1963.

12. Hunter, J.S. and T.H. Naylor, "Experimental Designs for Computer Simulation Experiments," Management Science, vol. 16, no. 7, pp.422-434, March 19 1970.

13. Ito, K., "On the Effect of Heteroscedasticity and Nonnormality upon some Multivariate Test Procedures," In P.R. Krishnaiah (ed.), Multivariate Analysis II, Academic Press, New York, 1969, pp.87-120.

14. ________ and W.J. Schull, "On the Robustness of the T^2 Test in Multivariate Analysis of Variance When Variance-covariance Matrices are not Equal," Biometrika, vol.51, Parts 1 and 2, pp.71-82, June 1964.

15. Kleijnen, Jack P.C. Statistical Techniques in Simulation, Parts I and II, Marcel Dekker, New York, 1975.

16. ________, "Analysis of Simulation with Common Random Numbers," Simuletter, vol.11, no.2, pp.7-10, Winter 1980.

17. Kotler, P. "A Guide to Gathering Expert Estimates," Business Horizons, vol. 13, no.5, pp.79-87, October 1970.

18. McArdle, J.J. "The Statistical Analyses of Monte Carlo Simulation Data Using the Techniques of Discrete Multivariate Analysis,"Proceedings, Computer Science and Statistics: 10th Annual Symposium on the Interface, April 14-15, 1977, pp.241-246.

19. Mardia, K.V. "Assessment of Multinormality and the Robustness of Hotelling's T^2 Test," Applied Statistics, vol.24, no.2, pp.163-171, 1975.

20. Mihram, G.A. Simulation: Statistical Foundations and Methodology, Academic Press, New York, 1972.

21. Miller, R.G., Jr. Simultaneous Statistical Inference, Springer-Verlag, New York, 1981.

22. Montgomery, D.C.and V.M. Bettencourt, "Multiple Response Surface Methods in Computer Simulation," Simulation vol. 29, no. 4, pp.113-121, Oct.1977.

23. Morrison, D.F. Multivariate Statistical Methods, McGraw-Hill, N.Y., 1976.

24. Naylor, T.H., J.L. Balintfy, D.S.Burdick, and K.Chu. Computer Simulation Techniques, Wiley, New York, 1966.

25. ________, D.S. Burdick, and W.E. Sasser, "Computer Simulation Experiments with Economic Systems: The Problem of Experimental Design," Journal of ASA vol 62, no.320, pp.1315-1337, 1967.

26. Olson, C.L. "Comparative Robustness of Six Tests in Multivariate Analysis of Variance," Journal of the ASA, vol. 69, no. 348, pp.894-908, 1974.

27. Popoviciu,N. and Vaduva, I. "χ^2-Test of Goodness of Fit for Multivariate Normal Distribution,"Economic Computation and Economic Cybernetics Studies and Research, vol.13, no.3, pp.93-103, 1979.

28. Puri, M.L. and P.K. Sen. Nonparametric Methods in Multivariate Analysis, Wiley, New York, 1971.

29. Schruben, L.W. "Control of Initialization Bias in Multivariate Simulation Response," Communications of the ACM, vol.24,no.4,pp.246-252, 1981.

30. Seila, A.F. "Multivariate Estimation in Regenerative Simulation," Operations Research Letters, vol.1,no.4, pp.153-156, September 1982.

31. Shannon, R.E. "Simulation: A Survey with Research Suggestions," AIIE Transactions, vol. 7, no. 3, pp.289-301, September 1975,

32. ______. Systems Simulation: The Art and Science, Prentice-Hall, Englewood Cliffs, New Jersey, 1975.