Analysis of Multispecies Microcosm Experiments

by

Donald E. Mercante

Dissertation submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Statistics

APPROVED:

Eric P. Smith, Chairman

Klaus H. Hinkelmann

Marvin Lentner

Marion R. Reynolds, Jr.

Robert V. Foutz

March 19, 1990

Blacksburg, Virginia

ANALYSIS OF MULTISPECIES MICROCOSM EXPERIMENTS

by

Donald E. Mercante

Committee Chairman: Eric P. Smith
Statistics

(ABSTRACT)

Traditionally, single species toxicity tests have been the primary tool for assessment of hazard of toxic substances in aquatic ecosystems. These tests are inadequate for accurately reflecting the impact of toxicants on the community structure inherent in ecosystems. Multispecies microcosm experiments are gaining widespread acceptance as an important vehicle in understanding the nature and magnitude of effects for more complex systems.

Microcosm experiments are complex and costly to conduct. Consequently, sample sizes are typically small (8–20 replicates). In addition, these experiments are difficult to analyze due to their multivariate and repeated measures nature.

Working under the constraint of small sample sizes, we develop inferential as well as diagnostic methods that detect and measure community changes as a result of an intervention (i.e. toxicant), and assess the importance of individual species.

A multi-factorial simulation analysis is used to compare several methods. The Multi-Response Permutation Procedure (MRPP) and a regression method incorporating a correlation structure are found to be the most powerful procedures for detecting treatment differences.

The MRPP is particularly suited to experiments with replication and when the response variable may not be normally distributed. The regression model for dissimilarity data has the advantage of enabling direct estimation of many parameters not possible with the MRPP as well as the magnitude of treatment effects.

A stepwise dependent variable selection algorithm with a selection criterion based on a conditional p-value argument is proposed and applied to a real data set. It is seen to have advantages over other methods for assessing species importance.

To My Wife, Teresita

Acknowledgements

I would sincerely like to thank my advisor Dr. Eric P. Smith for his patience, diligent support and guidance in my academic pursuits at V.P.I.&S.U. It is easy to lose one's way in early research, and he was careful to guide me back on course many times. I would like to thank my graduate committee members for there willingness to serve on my committee and give so freely there time and assistance to help in the completion of this dissertation. The committee members are Prof. Klaus H. Hinkelmann, Prof. Marvin Lentner, Dr. Marion R. Reynolds, Jr. and Dr. Robert V. Foutz.

Perhaps the most fond memories from graduate school are of the many friends I made during my stay in Blacksburg. It would have been difficult to have progressed this far without there help, support and encouragement. To many I owe thanks, particularly to Paul T. Savarese, Barbara Kuzmak, Philip J. Ramsey and Rich L. Einsporn.

When I look back on the long journey it was to get this far, the one thing I am instantly reminded of is how important family support is and how much mine has met

to me. It would have been impossible to finish, let alone begin, without the support, tremendous patience, understanding and love, of my wife, Teresita. Together with our two beautiful daughters, Lorenza Maria and Anna Alicia, we endured, persevered and through love and understanding became a much kinder and gentler family. To their innumerable sacrifices I owe an eternal debt.

I would like to extend special thanks to my parents, Mr. and Mrs. Ignatius Dominick Mercante, and parents-in-law, Mr. and Mrs. Jacob F. Musacchia, Sr., for their everlasting enthusiasm and support. They have given so much and have never asked for anything in return. Special thanks to Professor Miguel Guzman, a mentor and friend. Finally, I wish to thank Bob and Emily Stuart for their gracious hospitality and friendship to me and my family.

Table of Contents

List of Tables

List of Illustrations

CHAPTER I

INTRODUCTION

§ 1.1 *INTRODUCTION TO MULTISPECIES MICROCOSM EXPERIMENTS*

Traditionally, the single species (univariate) toxicity test has been the primary tool for assessment of hazard of toxic substances in aquatic ecosystems. These tests, however, often do not reflect the impact of toxic substances on the community structure of ecosystems. Multispecies microcosm experiments play an increasingly important role in understanding the nature and magnitude of effects on more complex systems (e.g., Taub 1976, Giesy 1980, NRC 1981, Cairns 1985, Cairns 1986). Because these tests incorporate properties of ecosystems (i.e., community structure) that cannot be studied in exposure of single species to stress, they may provide an additional basis for prediction of the ecological effects of toxic exposure and improve the accuracy of evaluations of hazard.

While there is no typical or standardized multispecies toxicity test, many studies share a number of common features and objectives. In many microcosm experiments a researcher has a number of doses of a toxicant that are of interest. These are then

applied to the test systems and the communities are measured in some way over time. Phytoplankton, zooplankton, benthic invertebrate and fish communities are the types of communities most often studied. For example, Dewey (1986) in a recent study on the effects of atrazine on benthic insect communities applied concentrations of 0, 20, 100 and 500 μg/l to eight experimental ponds. Samples of the communities were made at a number of time periods over five months. In these studies there are typically two factors, the treatment or toxicant and the time factor. Furthermore, the studies are really multivariate studies with the species as the variables. When the treatments are studied over time, the experiments often should be analyzed as a repeated measures design.

The usual analyses of these experiments do not account for the multivariate nature of the data nor the repeated measurements. Separate univariate analyses are conducted for each time point to avoid dealing with repeated measurements. Problems with multivariate data, however, remain improperly addressed. It is this problem and those associated with the inherently small samples sizes for multispecies toxicity tests which provide much of the motivation for this dissertation.

§ 1.2 *SCOPE OF DISSERTATION*

Much of the recent work in multispecies toxicity testing has been on non-inferential techniques such as Multi-Dimensional Scaling (MDS) and other graphical methods. Here we attempt to develop inferential as well as diagnostic methods that will detect and measure community changes as a result of an intervention (i.e., toxicant) and assess the role of individual species in those changes.

The dissertation is arranged into five chapters. The remainder of this chapter is devoted to major design issues arising in microcosm experiments. Several methods useful in the analysis of microcosm experiments are presented in Chapter II. Results of a power simulation study which provide insight into the utility of each method and identifies the best methods for further investigation are presented in detail. Additionally, the results of three smaller simulation studies using the permutation and regression methods are presented.

Chapters III gives consideration to the selection of variables in a multispecies experimental framework. Concepts relating to dependent variable (species) importance and selection are proposed and developed in the framework of the Multi-Response Permutation Procedure (MRPP) for experiments with replication. Estimation and testing in regression using permutation and parametric techniques are explored in Chapter IV. The dissertation concludes as Chapter V with a summary and concluding remarks regarding present and future research in this area.

§ 1.3 *DESIGN ISSUES*

Design issues will not be considered in extended detail in the dissertation. However, it would seem imprudent to undertake any involved treatment of analysis without considering the underlying design issues relevant to both single and multispecies experiments. The multiplicity of objectives commonly encountered in microcosm experiments leads to designs which are not optimal. For example, researchers may be interested in description of effects, inference about an effect due to a toxicant, joint effects of two toxicants, estimation of parameters such as the no observable effect level (NOEL). Only through careful planning and selection of experimental designs can these multiple objectives be properly addressed.

Issues paramount to these designs requiring careful attention are replication of treatments, sample size and power, optimality criteria in design selection, choice of number and spacing of dose levels, inference on "safe" dose and defining the dose-response curve. Completely randomized designs (CRD) are most often used, however, it may be of benefit to view these designs in the context of regression or response surface.

A central theme in improving any experimental design is the reduction of variation, particularly in the mean treatment response. Subsampling is a device employed by some researchers to increase precision. Subsampling is usually not advisable as it is not true replication but a repeated measurement on the same

experimental unit.

Let $\overline{Y}_{i..}$ represent the sample mean for treatment (dose) i averaged over replicates and subsamples, σ_ϵ^2 the error among experimental units treated alike, σ_η^2 the error incurred from sampling the same experimental unit repeatedly, n the number of subsamples and r the number of replicate experimental units per treatment level, then the variance of a treatment mean can be expressed as

$$\mathrm{Var}(\overline{Y}_{i..}) = \frac{\sigma_\eta^2 + n\sigma_\epsilon^2}{nr} = \frac{\sigma_\eta^2}{nr} + \frac{\sigma_\epsilon^2}{r}. \qquad [1.1]$$

Increased subsampling will reduce the overall variance of treatment means by reducing that part associated with subsampling error. The part associated with error among microcosms is unaffected. Usually $\sigma_\eta^2 \ll \sigma_\epsilon^2$ so that reducing the component associated with subsampling error by increasing n accomplishes little. Increasing treatment replications (r) results in a more dramatic reduction of variance since both components of error are simultaneously reduced.

Variance of treatment means when r replications per treatment and n observations per replication are taken can be compared with the variance obtained when r' replications and $n' = 1$ observation per experimental unit by considering the following inequality

$$\frac{\sigma_e^2}{r'} < \frac{\sigma_\eta^2 + n\sigma_\epsilon^2}{nr} \qquad [1.2]$$

where $\sigma_e^2 = \sigma_\eta^2 + \sigma_\epsilon^2$. Replication is more efficient whenever the above holds (Hinkelmann and Kempthorne, unpublished manuscript).

Accurate determination of sample size in microcosm experiments is difficult due to the multi-objective requirements usually imposed on the experimental design. Sample size should be sufficiently large to enable estimation of a "safe dose" or NOEL, provide adequate power for statistical tests and information on the dose response relationship. Factors affecting sample size are range and number of dose levels and the magnitude of experimental error. Unfortunately, due to the high cost of replication and the complexity of these experiments sample sizes have been very small. A typical experiment will involve 2-5 dose groups with 8-20 replications in total.

When the number of replicates is fixed, power of statistical tests can be greatly improved through proper design selection. Power depends on sample size, range of dose or treatment levels, error variance and unknown parameters. Sample size and dose range are controlled by the experimenter. Calculating power in a regression or response surface setting with one treatment factor at many levels involves the noncentrality parameter δ from a noncentral F distribution

$$\delta = \frac{\beta^2 \mathbf{X}'\mathbf{X}}{\sigma^2} \qquad [1.3]$$

where $\mathbf{X}$ is the design matrix containing columns for an intercept and dose levels. By fixing values of β and σ^2, maximum power can be achieved by maximizing $\mathbf{X}'\mathbf{X}$. This is accomplished by spreading two dose levels as far apart as possible and placing half

the observations at either end. Increasing the dose range may increase the power of statistical tests but may also inflate prediction variances in desirable locations if the true model deviates from linearity. Many researchers believe 2 dose levels are too few and suggest many more may be necessary (Brown 1978, Giddings 1986, 1979).

One statistical concern which leads to additional doses is the "true" model. Two dose models lead to good statistical properties only when the true model is linear. If the true model is exponential or quadratic a serious bias can result. To protect against bias, additional dose levels are required. Then one may test directly the linear regression against a quadratic model using a lack-of-fit test (Myers 1986).

Although designs which are balanced, equally spaced (possibly on a log scale) and equally replicated are common, they are often not the optimal designs as is indicated in the work done by many researchers, most notably by David and Arens (1959), Krewski (et al. 1986, 1984) and Krewski and Kovar (1982) . Evidence suggests that designs that adhere to the following guidelines are either optimal or near optimal. When only two dose levels are being considered, allocation of $70-80\%$ of the sample to the low dose and $20-30\%$ to the high dose often approaches optimality. Equal spacing and allocation ratios of 1:2:1 and 1:2:2:1 in experiments with 3 or 4 dose levels, respectively, yield nearly optimal designs.

Decisions regarding design selection must be based on some objective criteria. The aforementioned optimal sample allocations are based on optimality criteria. David and Arens (1959) chose as their criterion minimized mean squared error

(MSE). Krewski and collegues based their criterion on minimizing the virtual safe dose (VSD).

When prediction variance of response values is considered as a criterion for design selection we found equally replicated designs to be optimal or nearly optimal. Myers (1986) formulates prediction variance at a given data location, $\underline{X}_0$, as

$$\mathrm{VAR}(\hat{y}(\underline{x}_0)) = \sigma^2 \, \underline{x}_0'(\mathbf{X}'\mathbf{X})^{-1} \, \underline{x}_0 \qquad [1.4]$$

where X represents the design matrix and x_0 is $\underline{X}_0$ adjusted for its mean. When there is a single toxicant or treatment, this may be written as

$$\mathrm{VAR}(\hat{y}(\underline{x}_0)) = \sigma^2 \, \frac{(X_0 - \bar{x})^2}{SS_{xx}} \qquad [1.5]$$

where SS_{xx} is the sum of squares due to X.

Many other optimality criteria such as the alphabetic optimality criteria (A, D, E, G) can be used to compare and select designs (Kiefer and Wolfowitz 1959). The two most popular criteria being G-optimality where interest is in minimizing the maximum prediction variance over all points, and D-optimality in which one wishes to minimize the generalized variance $\left|\sigma^2(\mathbf{X}'\mathbf{X})^{-1}\right|$ or equivalently maximize $\left|\sigma^2(\mathbf{X}'\mathbf{X})\right|$ (Kiefer and Wolfowitz 1959).

As an example of the effect of the location of design points (dose levels), consider

the data in Table 1.1. Ten 12 replicate completely randomized designs with $2-6$ levels over a range of $0-100$ were chosen for comparisons using average prediction variance (aside from σ^2), maximum prediction variance (G-optimality) and maximum $\left|(\mathbf{X}'\mathbf{X})\right|$ (D-optimality) as optimality criteria. Summaries of comparisons involving the above criteria are presented in the form of percentages and percent increases over the optimal design, D3.

To illustrate how a criterion such as prediction variance may be used to assess competing designs, let us compare the first three designs from Table 1.1. Prediction variance for design D1, which has a simple allocation ratio of 1:2:2:1 and equally spaced levels, is never better than that for the optimal design D3 (Figure 1.1). Design D2 which also has an allocation ratio of 1:2:2:1 and more replicates "loaded" at the low end of the range exhibits smaller variance than design D3 in the range $0-40$. If inference at lower doses is desired then design D2 may be preferred. However, prediction variance for D2 quickly increases above that for design D3 in the range $40-100$.

Although prediction variance is a valuable criterion for design selection there is a more useful criterion when model misspecification is present. One should bear in mind that nearly all dose response models can be viewed as approximations to some "true" model. Bias, as well as prediction variance, should be taken into account. Mean squared error (MSE) of predicted values is a useful criterion for incorporating both prediction variance and bias and is computed as

Table 1.1. Comparison of four criteria for 10 completely randomized designs with 12 replicates each.

Design	Dose Levels	Sample Allocation	Percent [1]$PV_3<PV_i$	Percent increase in [2]APV	[3]Max PV	[4]GV
D1	0, 33, 66, 100	1:2:2:1	99	37	73	145
D2	0, 2, 40, 100	1:2:2:1	61	47	141	98
D3	0, 100	1:1	---	---	---	---
D4	0, 50, 100	1:1:1	100	13	25	50
D5	0, 33, 66, 100	1:1:1:1	100	20	40	80
D6	0,20,40,60,80,100	1:1:1:1:1:1	99	30	57	114
D7	0, 33, 66, 100	2:2:1:1	68	32	98	97
D8	0, 50, 100	1:2:1	99	25	50	100
D9	0,20,40,60,80,100	1:2:3:3:2:1	100	57	113	226
D10	0, 100	9:3	63	33	100	33

[1] PV_i denotes prediction variance for design D_i

[2] APV denotes average prediction variance

[3] Max PV denotes maximum prediction variance

[4] GV denotes generalized variance

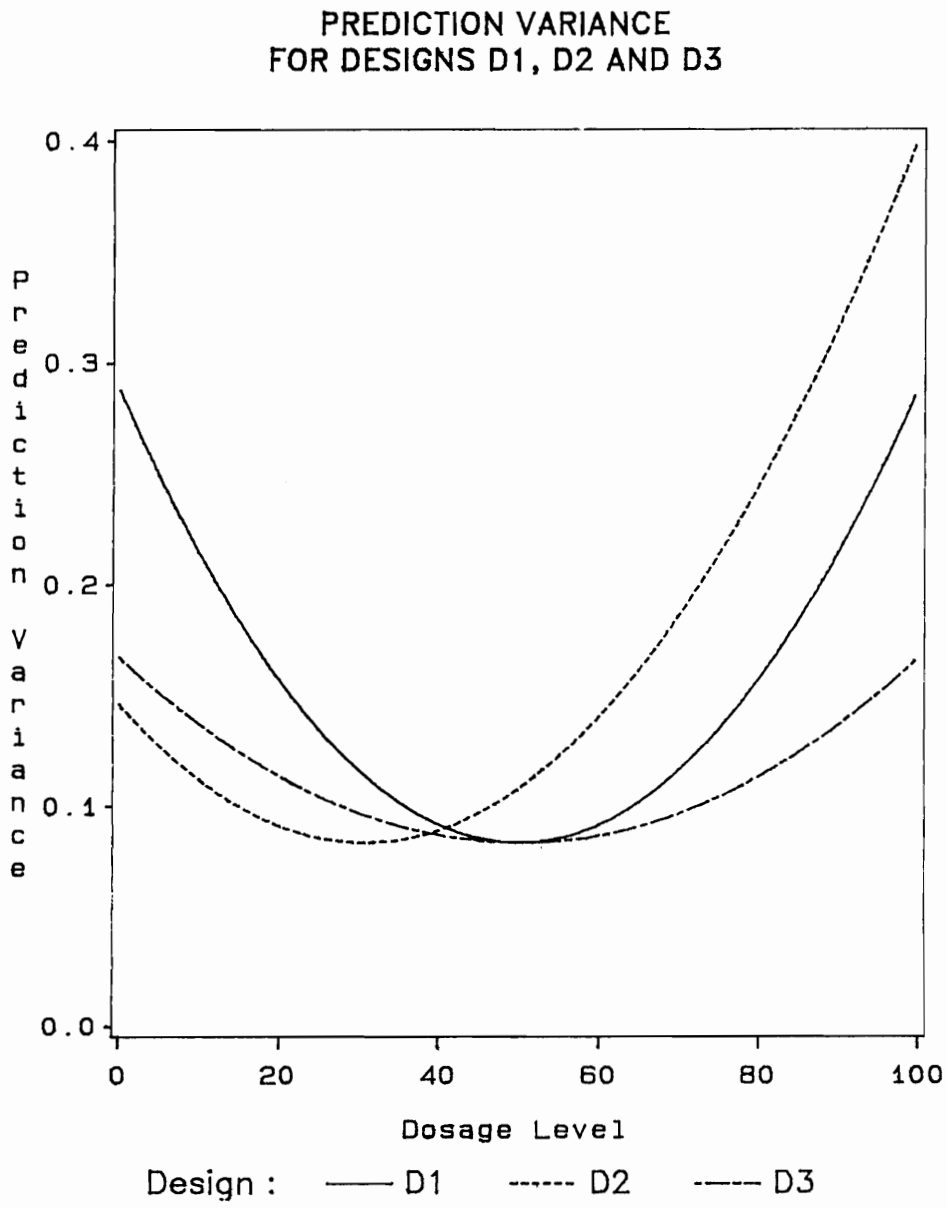

Figure 1.1. Prediction variance versus dose for designs D1, D2 and D3 from Table 1.1.

$$\frac{\text{MSE}(\hat{y}_i)}{\sigma^2} = \frac{\text{VAR}(\hat{y}_i)}{\sigma^2} + \frac{(\text{Bias}(\hat{y}_i))^2}{\sigma^2} \quad \text{(Myers 1986).} \qquad [1.6]$$

A plot of MSE versus dose level for three designs described in Table 1.1 appears in Figure 1.2. The model under consideration is quadratic in the dose variable rather than linear. Note that the design which results in the best power for the linear model (D3) has the worst MSE for most of the dose range. When interest is in very low doses (e.g. estimation of NOEL) or high doses (e.g. estimation of MTD) design D3 remains the best design. However, if interest centers in the middle of the dose range (e.g. estimation of LC_{50}) design D1 is the best design when using the MSE criterion.

It is evident that an universally optimal design does not exist. Therefore an obvious approach to design selection is to compromise. The design chosen should have good power and be able to estimate key parameters such as a NOEL value with reasonably good accuracy. Such a design may have three or four doses near the hypothesized NOEL level and one or two doses more extreme to increase power and precision.

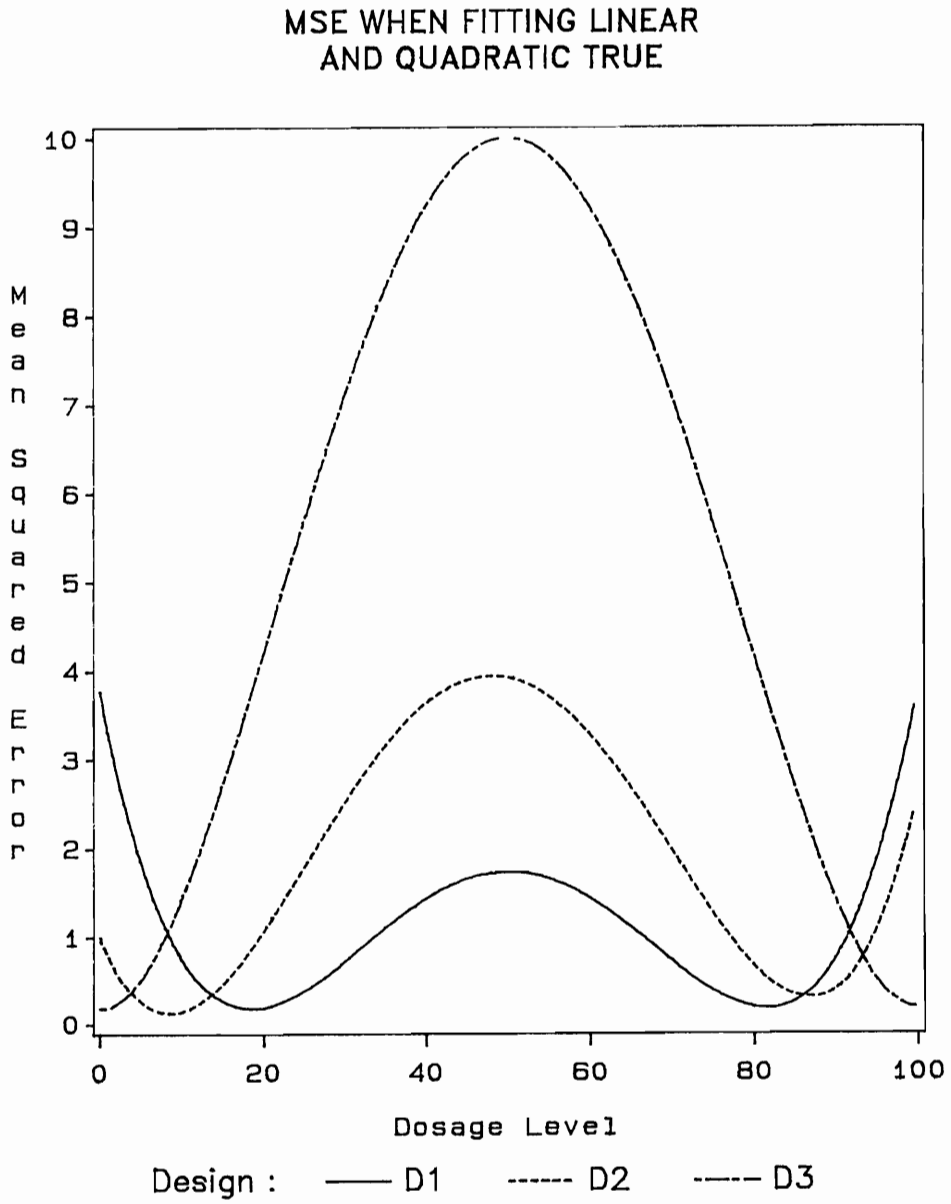

Figure 1.2. Graph of MSE against dose for designs D1, D2 and D3 from Table 1.1 when fitting a linear model and the true model is quadratic.

CHAPTER II

METHODS FOR ASSESSING COMMUNITY CHANGE

§ 2.1 *INTRODUCTION*

In analyzing multispecies data, there are difficulties applying standard statistical techniques. While the data is multivariate, standard techniques often may not be applied. For example, if there are 100 species (variables) that were identified by the researcher in the study, one needs at least 100 replicates to make inferences about changes in the community using inferential multivariate methods. What is commonly done is to (1) focus on univariate techniques or (2) use some summary measure of community composition, such as diversity or (3) use noninferential techniques, such as principal components analysis or cluster analysis. A further problem is that methods based on the normal distribution may not be applicable because of the large number of zeros in the data sets.

Following are brief descriptions and development of analytic methods which may

be or have been utilized in analyses of data from microcosm experiments. Parametric ANOVA and regression based approaches are presented, followed by material on permutation ANOVA and regression techniques requiring fewer assumptions for proper inference.

§ 2.2 *METHODS OF COMMUNITY ASSESSMENT*

§ 2.2.1 ANOVA Approach

Analysis of variance approaches are quite appealling intuitively to most researchers. There are several approaches possible for analyzing this type of data that are related to the ideas of the analysis of variance and tests based on normal theory which are multivariate in nature. For simplicity we assume throughout there are only two treatments. Also, the data are assumed to be species abundances (or biomass). Thus, the data may be viewed as vectors of abundances; $\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_{n_1}, \underline{X}_{n_1+1}, \ldots, \underline{X}_{n_1+n_2}$, where n_1 and n_2 are the respective sample sizes for groups 1 and 2. Also, $\underline{X}_i = (X_{i1}, X_{i2}, \ldots, X_{iP})$, where P is the number of species in the experiment. It is also assumed that the total abundance for each replicate is the same. This is an important assumption because the bias of an estimated measure of similarity (see below) is related to the total abundance (Smith and Zaret 1982, Ricklefs and Lau 1980). This assumption is realistic as standard methods specify a fixed number of organisms.

A number of researchers have recently suggested methods for the inferential analysis of multispecies data using a measure of similarity (Hruby 1987, Boyle et al. 1984, Dyer 1978). These methods have as a first step, the summarization of data using some measure of similarity. The type of measure depends in part on the type of data. Two popular measures that are based on abundances or proportion of

abundances are the proportional similarity measure between replicates i and j, given by

$$PS_{ij} = \sum_{k=1}^{P} \min(X_{ik}, X_{jk}), \qquad [2.1]$$

and Stander's (1970) or the cosine measure

$$COS_{ij} = \frac{\sum X_{ik} X_{jk}}{\sqrt{\sum X_{ik}^2 \sum X_{jk}^2}}. \qquad [2.2]$$

All sums are from k=1 to P.

A widely used measure of dissimilarity is Euclidean distance

$$D_{ij} = \sqrt{\sum_{k=1}^{P} (X_{ik} - X_{jk})^2}. \qquad [2.3]$$

Table 2.1 gives a matrix of dissimilarities using the Euclidean distance index of dissimilarity. This matrix is of size N×N, where N is the total number of replicates. Whenever a difference exists between the two treatment groups there will be a natural partition of the dissimilarity matrix into 2 components, the within and between. Between refers to the dissimilarity between replicates from different treatments and within refers to the dissimilarity within the same treatment group. In

Table 2.1. Hypothetical data on 5 species from 5 replicates at 2 locations. Dissimilarities using Euclidean distance given below the matrix of abundances.

Species

Site	Rep	1	2	3	4	5
1	1	6	24	6	18	0
1	2	11	18	0	16	0
1	3	9	26	0	13	4
1	4	11	13	7	12	4
1	5	7	23	6	17	6
2	6	3	14	0	12	5
2	7	8	8	0	10	5
2	8	6	16	0	7	9
2	9	8	3	0	6	9
2	10	4	13	0	11	11

Reps

Reps	1	2	3	4	5	6	7	8	9	10
1	0	9.9	7.2	12.4	4.2	13.0	18.2	12.2	23.3	15.5
2	9.9	0	8.3	8.8	9.2	9.8	11.6	8.1	16.8	11.7
3	7.2	8.3	0	14.9	7.2	13.7	18.4	11.6	23.8	15.6
4	12.4	8.8	14.9	0	10.8	10.9	9.6	10.0	13.5	11.6
5	4.2	9.2	7.2	10.8	0	11.6	16.3	9.8	21.3	13.0
6	13.0	9.8	13.7	10.9	11.6	0	7.8	4.1	12.4	4.2
7	18.2	11.6	18.4	9.6	16.3	7.8	0	8.3	5.3	7.0
8	12.2	8.1	11.6	10.0	9.8	4.1	8.3	0	13.1	4.1
9	23.3	16.8	23.8	13.5	21.3	12.4	5.3	13.1	0	10.8
10	15.5	11.7	15.6	11.6	13.0	4.2	7.0	4.1	10.8	0

the example, N = 10 with 5 replicates of each treatment. Note that while there are a total of only 10 replicates, there are 20 estimates of within dissimilarity and 25 estimates of between dissimilarity.

The second step is to treat the similarities/dissimilarities as data and analyze this data using analysis of variance techniques (Boyle et al. 1984, Hruby 1987, Brock 1977 and others). For example, in the two group case, Boyle et al. recommend testing for differences between the between and within mean similarities using a two sample independent t-test (treating the within similarities and the between similarities as two independent groups of observations). Let S_{ij} denote a general similarity measure with i and j equal to 1, 2, ..., n_1+n_2. One would use

$$t_B = \frac{\overline{W} - \overline{B}}{S_P \sqrt{\frac{1}{m_W} + \frac{1}{m_B}}} \qquad [2.4]$$

where m_W and m_B are the number of within and between similarities used to compute the means. The pooled variance estimate, S_P^2, is given by

$$S_P^2 = \frac{\sum_{i,j\ within} (S_{ij} - \overline{W})^2 + \sum_{k,l\ between} (S_{kl} - \overline{B})^2}{(m_W + m_B - 2)}. \qquad [2.5]$$

$\overline{W}$ and $\overline{B}$ represent the mean within and between similarities, i.e.,

$$\overline{W} = \frac{\sum_{i>j}^{n_1} S_{ij} + \sum_{k>l>n_1}^{n_1+n_2} S_{kl}}{\frac{n_1(n_1-1)}{2} + \frac{n_2(n_2-1)}{2}} = \frac{\sum_{i>j}^{n_1} S_{ij} + \sum_{k>l>n_1}^{n_1+n_2} S_{kl}}{m_W} \qquad [2.6]$$

and

$$\overline{B} = \frac{\sum_{i=1}^{n_1} \sum_{j=n_1+1}^{n_1+n_2} S_{ij}}{n_1 n_2} = \frac{\sum_{i=1}^{n_1} \sum_{j=n_1+1}^{n_1+n_2} S_{ij}}{m_B} . \qquad [2.7]$$

Note that *i,j within* refers to subscripts for the within similarities, while *k,l between* refers to subscripts for the between similarities. The null hypothesis of no group differences is rejected if the t-statistic exceeds the critical value from a t-distribution with $m_W + m_B - 2$ degrees of freedom.

While Boyle's method is appealing, the measures of similarity are not independent (e.g., S_{12} and S_{13} have one sample in common). For the data in Table 2.1, 45 "data points" (dissimilarities) are used to estimate the difference between the within and between dissimilarity while there are only 10 true replicates. Therefore, the method cannot be recommended for use in general.

The statistic t_B is a Mantel-Valand statistic, or more generally a Mantel statistic (Mantel and Valand 1970, Mantel 1967). Mantel and Valand suggest using a standard normal approximation for inference, however, several authors have indicated that this approximation is not valid, even for large samples (Mielke 1979). Mielke (1986 and references therein) suggests that methods based on randomization or permutation approaches, such as the Multi-Response Permutation Procedure (MRPP), are the best approaches for inference for this type of problem. Mielke has also shown that if the sample sizes differ, then $\overline{W} - \overline{B}$ is an inefficient estimator (other estimators have smaller variance). The criticisms are for statistics of this type in

general. The case that is dealt with here is special in that the distances or similarities themselves may be approximately normal if the number of individuals sampled are used (not number of replicates) is reasonably large.

§ 2.2.2 Regression Approach - Dyer's Method

Dyer (1978) proposed a parametric regression approach for analyzing similarity/dissimilarity data based on normal theory. His model is appropriate for comparing two or more multispecies (multivariate) samples. Multiple independent (environmental) variables can be included in the model which is given as

$$D_{ij} = \beta_0 + \beta_1 \delta_{ij}^1 + \beta_2 \delta_{ij}^2 + \cdots + \beta_m \delta_{ij}^m + \epsilon_{ij} \qquad [2.8]$$

where

D_{ij} is the dissimilarity of samples i and j,

δ_{ij}^l is a known function of i and j corresponding to the l^{th} independent variable,

β_l represents the contribution of the l^{th} independent variable to total dissimilarity,

ϵ_{ij} is an error term with expectation of zero and

$\mathrm{Var}(\epsilon_{ij}) = \sigma^2$ for i, j distinct,

$\mathrm{Cov}(\epsilon_{ij}, \epsilon_{jk}) = \rho$ for i, j, k distinct,

$\mathrm{Cov}(\epsilon_{ij}, \epsilon_{kh}) = 0$ for i, j, k, h distinct.

Computing estimates for model parameters Dyer presents OLS estimates for β as

$$\hat{\beta} = (X'X)^{-1}X'Y. \qquad [2.9]$$

In a preliminary step to obtain unbiased estimates for model parameters σ^2 and ρ,

the sample variance and covariance are defined as

$$\sigma_s^2 = \frac{2}{n(n-1)}[\mathbf{Y}'\mathbf{Y} - tr(\hat{\beta}\hat{\beta}'\mathbf{X}'\mathbf{X})] \qquad [2.10]$$

$$\rho_s = \frac{1}{n(n-1)(n-2)}[\mathbf{Y}'\mathbf{C}\mathbf{Y} - tr(\hat{\beta}\hat{\beta}'\mathbf{X}'\mathbf{C}\mathbf{X})] \qquad [2.11]$$

which has expectation

$$E(\sigma_s^2) = \sigma^2 - \frac{2}{n(n-1)}[(m+1)\sigma^2 + tr((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{C}\mathbf{X})\rho] \qquad [2.12]$$

$$E(\rho_s) = \rho - \frac{1}{n(n-1)(n-2)}[tr((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{C}\mathbf{X})\sigma^2 + tr((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{C}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{C}\mathbf{X})\rho] \qquad [2.13]$$

Where C is a $n(n-1)/2 \times n(n-1)/2$ matrix corresponding to the covariances among the dissimilarity values. Entries of C are either 0 or 1. The j, k^{th} entry of C is 1 if the j^{th} and k^{th} dissimilarity values have exactly one sample in common, e.g., D_{12} and D_{24} have sample 2 in common.

By substituting ρ_s for $E(\rho_s)$ and σ_s^2 for $E(\sigma_s^2)$, estimates are obtainable for the model parameters ρ and σ^2. An unbiased estimate for the variance-covariance matrix of $\hat{\beta}$ is given as

$$\widehat{Var}(\hat{\beta}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{C}\hat{\rho} + \mathbf{I}\hat{\sigma}^2)\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}. \qquad [2.14]$$

Dyer presented ideas relating only to estimation which are approximate and based on normal theory. Tests of the coefficients can be made by constructing t-tests of the form

$$t_D = \frac{\hat{\beta}_i}{S_{\hat{\beta}_i}} \qquad [2.15]$$

where $S_{\hat{\beta}_i}$ is the i^{th} diagonal element of $\widehat{Var}(\hat{\beta}_i)$.

Note that if C=I, $\rho = 0$ and δ = grouping variable, then the regression (Dyer's) approach can be made to yield the same results as the ANOVA (Boyle's) approach. Hence, Dyer's regression method can be considered a generalization of the ANOVA approach and it becomes apparent the real utility of this approach is in its generality and ability to handle a wide array of models.

§ 2.2.3 Hotelling T^2 and the Behrens-Fisher Problem

Hotelling's T^2 statistic is widely used for testing the hypothesis $H_0: \mu_1 = \mu_2$ versus $H_1: \mu_1 \neq \mu_2$ when two random samples are obtained and the random variates follow multivariate normal distributions. It is assumed that the variates have common, yet unknown, full rank covariance matrix Σ. Details of Hotelling's T^2 statistic can be found in most multivariate texts (e.g., Morrison 1976, Anderson 1958). The T^2 statistic is the multivariate analog of the univariate t-test statistic. It can be developed by the union-intersection principle or maximum likelihood and is given as

$$T^2 = \frac{N_1 N_2}{N_1 + N_2} (\overline{X}_1 - \overline{X}_2)' S^{-1} (\overline{X}_1 - \overline{X}_2) \sim T^2_{\alpha;\, P,\, N_1,\, N_2 - P - 1} \qquad [2.16]$$

which can be transformed into an F statistic by

$$F = \frac{N_1 + N_2 - P - 1}{(N_1 + N_2 - 2)P} T^2 \sim F_{\alpha;\, P,\, N_1 + N_2 - P - 1} \qquad [2.17]$$

where $\overline{X}_1$ and $\overline{X}_2$ are the sample mean vectors for variates 1 and 2, respectively, and S is the pooled sample covariance matrix computed as

$$S = \frac{1}{N - P} \sum_{k=1}^{P} \sum_{i=1}^{N_k} (X_{ik} - \overline{X}_k)(X_{ik} - \overline{X}_k)' \qquad [2.18]$$

When $\Sigma_1 \neq \Sigma_2$ we have the multivariate counterpart to the Behrens-Fisher

problem. We can use the multivariate extension of the Welsh test statistic (Seber 1984, Johnson and Wichern 1982) given as

$$T_{B\text{-}F} = (\overline{X}_1 - \overline{X}_2)' \left[\frac{1}{N_1} S_1 + \frac{1}{N_2} S_2 \right]^{-1} (\overline{X}_1 - \overline{X}_2) \approx \chi^2_P \qquad [2.19]$$

to test for group differences.

§ 2.2.4 Multi-Response Permutation Procedures (MRPP)

An alternate procedure which is valid for this data is a permutation or randomization test. The permutation test is carried out by switching data from replicates of different treatments and noting changes in the within and between similarities (dissimilarities). Under the null hypothesis, switching should have no effect on the within and between similarities. Under the alternate hypothesis, switching should reduce the within similarity and increase the between similarity (see Table 2.1 for an example). If a large number of switches are made, a test can be performed.

Multi-response permutation procedures have been studied and described in detail (Mielke 1986, 1979, 1978, 1976, Mielke et al. 1976, O'Reilly and Mielke 1980, Berry and Mielke 1984, Brockwell et al. 1982, Robinson 1983, Tracy and Tajuddin 1986, 1985). MRPP are appropriate for analyses involving multiple response variables which are summarized using a metric measure, usually a symmetric distance measure such as Euclidean distance. They avoid typical distributional assumption problems associated with most parametric tests and are valid for ordinal and higher scaled data. Exact permutation procedures applicable to multi-response data have been derived as well as approximations using the beta and normal distributions.

In the terminology of Mielke (1976) consider a finite population of N elements $\Omega=(\omega_1,\ldots,\ \omega_N)$ which are partitioned , *a priori*, into g+1 mutually exclusive and

exhaustive groups $S_1, \ldots, S_g, S_{g+1}$ (note that S_{g+1} may be empty). Associated with any ω_I may be P response measurements, say, $X_I=(X_{1I}, \ldots, X_{PI})$. Let $n_i \geq 2$ be the sample size for group S_i $(i=1, \ldots, g)$, $K=\sum_{i=1}^{g} n_i$ and $n_{g+1}=N-K\geq 0$.

Each pair of P-dimensional response vectors is summarized by a (symmetric) distance function such as

$$\Delta_{IJ} = \left\|X_I - X_J\right\| = \left[\sum_{K=1}^{P} [X_{KI} - X_{KJ}]^2\right]^{\frac{v}{2}} \qquad [2.20]$$

When v=1, Δ_{IJ} is Euclidean distance (L_2 norm). The test statistic for the MRPP is given by

$$\delta = \sum_{i=1}^{g} C_i f_i \qquad [2.21]$$

where

$$f_i = \binom{n_i}{2}^{-1} \sum_{I<J} \Delta_{IJ} I_{S_i}(\omega_I) I_{S_i}(\omega_J),$$

$$C_i = \frac{n_i}{K} \quad \text{for } i=1, \ldots, g$$

$I_{S_i}(\omega_I) = 1$ if $\omega_I \in S_i$ and $I_{S_i}(\omega_I) = 0$ if $\omega_I \notin S_i$, for $1 \leq I \leq J \leq N$.

The null hypothesis of no treatment effect can be tested using δ. Under the null hypothesis each of the

$$M = \frac{N!}{\prod_{i=1}^{g+1} n_i!} \qquad [2.22]$$

possible outcomes is equally probable. If $C_i = \frac{(n_i - 1)}{(K-g)}$ and v=2, then $\delta = \bar{B} - \bar{W}$ from the ANOVA approach and it can be shown that all the methods give the same estimates. The difference lies in what is being tested.

Mielke (1976, 1979) and O'Reilly and Mielke (1980) describe methods for employing approximate tests (based on the beta and normal distributions) for the case when M is impractically large. One such approximate test could be based on the first two moments of permutation distribution for δ, given as

$$\mu_\delta = \binom{N}{2}^{-1} \sum_{I<J} \Delta_{IJ} \qquad [2.23]$$

$$\sigma_\delta^2 = 2\left(\left(\sum_{i=1}^{g} n_i^{(2)}\right)^{-1} - (N^{(2)})^{-1}\right)\bullet\left(\binom{N}{2}^{-1} \sum_{I<J} \Delta_{IJ}^2 - (N^{(4)})^{-1} \sum_{I,J,K,L} \Delta_{IJ}\Delta_{KL}\right) +$$

$$4\left(\left(\sum_{i=1}^{g} n_i^{(2)}\right)^{-2} \sum_{i=1}^{g} n_i^{(3)} - (N^{(2)})^{-1}(N-2)\right)\bullet$$

$$\left((N^{(3)})^{-1} \sum_{I,J,K} \Delta_{IJ}\Delta_{IK} - (N^{(4)})^{-1} \sum_{I,J,K,L} \Delta_{IJ}\Delta_{KL}\right) \qquad [2.24]$$

where, $N^{(j)} = \frac{N!}{(N-j)!}$.

A test statistic based on the standard normal distribution could be constructed as

$$Z = \frac{\delta - \mu_\delta}{\sigma_\delta} \qquad [2.25]$$

When enumeration of all possible permutations is feasible, algorithm *AS* 179 (Berry 1982) can be used to find and process each permutation.

A pronounced negative skewness of the underlying permutation distribution often exists and hence leads to situations where the null distribution of the MRPP test statistic is non-normal. Mielke (1976, 1979) uses the first three moments of the permutation distribution of the MRPP to arrive at a better approximate test based on the beta distribution.

Let $K_3(\delta)$ denote the third cumulant of δ, then the skewness of δ can be expressed as

$$\gamma_\delta = \frac{K_3(\delta)}{\sigma_\delta^3} \qquad [2.26]$$

where

$$K_3(\delta) = E(\delta^3) - 3\mu_\delta\sigma_\delta^2 - \mu_\delta^3 \, . \qquad [2.26]$$

To calculate γ_δ, it is necessary to determine μ_δ, σ_δ^2 and $E(\delta^3)$. The first two quantities are given above, thus it is necessary only to determine $E(\delta^3)$ which is given as

$$E(\delta^3) = \left(\sum_{i=1}^{g} n_i^{(2)}\right)^{-3}\left\{4\sum_{i=1}^{g} n_i^{(2)} D(3)+8\sum_{i=1}^{g} n_i^{(3)}[3\,D(3')+D(3*)]\right.$$

$$+ 8\sum_{i=1}^{g} n_i^{(4)}[3\,D(3**)+D(3***)]+6\left(\sum_{i=1}^{g} n_i^{(4)}+2\sum_{i<i'} n_i^{(2)}n_{i'}^{(2)}\right)D(3'')$$

$$+ 12\left(\sum_{i=1}^{g} n_i^{(5)}+\sum_{i<i'}\left(n_i^{(3)}n_{i'}^{(2)}+ n_i^{(2)}n_{i'}^{(3)}\right)\right) D(3''') + \left(\sum_{i=1}^{g} n_i^{(6)}\right.$$

$$\left.\left.+ 3\sum_{i<i'}\left(n_i^{(4)}n_{i'}^{(2)}+ n_i^{(2)}n_{i'}^{(4)}\right)+ 6\sum_{i<i'<i''} n_i^{(2)}n_{i'}^{(2)}\, n_{i''}^{(2)}\right)D(3'''')\right\} \qquad [2.26]$$

where we define $D(\cdot)$ as follows. Let

$$d_{KJ}=\sum_{J'=1}^{N}\Delta^{K}_{J,J'} \quad \text{and} \quad d_K=\sum_{J=1}^{N} d_{KJ} \text{ for K=1, 2 and 3.}$$

Then

$$D(1)=\frac{1}{N^{(2)}}d_1, \quad D(2)=\frac{1}{N^{(2)}}d_2, \quad D(2')=\frac{1}{N^{(3)}}\left(\sum_{J=1}^{N} d_{1J}^2-d_2\right),$$

$$D(2'')=\frac{1}{N^{(4)}}\left(d_1^2-4\,N^{(3)}D(2')-2\,d_2\right), \quad D(3)=\frac{1}{N^{(2)}}\, d_3,$$

$$D(3')=\frac{1}{N^{(3)}}\left(\sum_{J=1}^{N} d_{1J}d_{2J}-d_3\right), \quad D(3'')=\frac{1}{N^{(4)}}\left(d_1d_2-4N^{(3)}D(3')-2d_3\right),$$

$$D(3*)=\frac{6}{N^{(3)}}\sum_{J_1<J_2<J_3}\Delta_{J_1,J_2}\Delta_{J_1,J_3}\Delta_{J_2,J_3},$$

$$D(3**)=\frac{6}{N^{(4)}}\left(2\sum_{J_1<J_2}\Delta_{J_1,J_2}d_{1J_1,1J_2}-2N^{(3)}D(3')-N^{(3)}D(3*)-d_3\right),$$

$$D(3***)=\frac{1}{N^{(4)}}\left(\sum_{J=1}^{N} d_{1J}^3-3N^{(3)}D(3')-d_3\right),$$

$$D(3''')=\frac{1}{N^{(5)}}\left(N^{(3)}d_1D(2')-4N^{(4)}D(3**)-2N^{(4)}D(3***)-4N^{(3)}D(3')-2N^{(3)}D(3*)\right),$$

$$D(3'''')=\frac{1}{N^{(6)}}\left(N^{(4)}d_1D(2'')-8N^{(5)}D(3''')-4N^{(4)}D(3'')-8N^{(4)}D(3**)\right).$$

An approximation to the permutational distribution of δ based on the first three moments was initially developed by Mielke (1976). Standardizing δ we get

$$t=\frac{(\delta-\mu_\delta)}{\sigma_\delta} \qquad [2.26]$$

which can be used in approximating the distribution of

$$t_\beta=t\left[\frac{\alpha\beta}{(\alpha+\beta)^2}(\alpha+\beta+1)\right]^{\frac{1}{2}}+\frac{\alpha}{(\alpha+\beta)} \qquad [2.26]$$

using a beta distribution with density given as

$$f(x)=\begin{cases} x^{\alpha-1}(1-x)^{\beta-1}/\ B(\alpha,\beta), & 0<x<1 \\ 0, \text{ otherwise} \end{cases} \qquad [2.26]$$

where α, $\beta>0$ and

$$\gamma_{\delta}=\left[2\frac{(\beta-\alpha)}{(\alpha+\beta+2)}\right]\left[\frac{(\alpha+\beta+1)}{\alpha\beta}\right]^{\frac{1}{2}}. \qquad [2.26]$$

Tracy and Tajuddin (1985) suggest the use of the fourth order moment to improve inference in certain situations. It is, however, quite cumbersome to employ as it utilizes 32 additional symmetric functions of the distance function.

§ 2.2.5 Mantel-Valand Test

Mantel (1967) and Mantel and Valand (1970) have developed a permutation procedure based on ranks which is conceptually the same as the MRPP. Their test statistic can be written as

$$Z = 2\sum_{i<j} X_{ij} R_{ij} \qquad [2.26]$$

where R_{ij} is a distance or similarity measure based on the ranks of the observations, and $X_{ij}=1$ when individuals i and j belong to the same group, and 0 otherwise. Z can be evaluated relative to its permutation distribution. For large values of M, Mantel and Valand report that Z is in the form of a Hoeffding *U* statistic and that it is approximately normally distributed. They propose a t-test based on Z's permutational expectation and standard deviation given as

$$t = \frac{Z - E(Z)}{\sigma(Z)} \qquad [2.27]$$

where E(Z) and $\sigma^2(Z)$ are defined as

$$E(Z) = \sum_{i \neq j} X_{ij} E(R_{ij}) = \sum_{i \neq j} X_{ij} \sum_{i \neq j} \frac{R_{ij}}{n(n-1)} \qquad [2.28]$$

and

$$\sigma^2(Z)=\mathrm{Var}(Z)=\sum_{\substack{i\neq j\\ k\neq l}} X_{ij}X_{kl}\,\mathrm{Cov}(R_{ij},R_{kl})=\sum_{\substack{i\neq j\\ k\neq l}} X_{ij}X_{kl}E(Y_{ij}Y_{kl})-\frac{\left(\sum_{i\neq j} X_{ij}\right)^2\left(\sum_{k\neq l} R_{kl}\right)^2}{n^2(n-1)^2} \qquad [2.29]$$

Mielke (1978) points out that the distribution of the Mantel-Valand test statistic is skewed and thus the statistic is not normally distributed, hence inferences based on this test may be incorrect. Correct inferences can be obtained by using Mielke's (1978) skewness correction.

When interest is in comparing mean similarities for the different treatment levels, a multiple comparisons randomization test developed by Foutz et al. (1985) can be used which controls the experimentwise error rate.

§ 2.3 *SIMULATION STUDIES*

§ 2.3.1 Introduction

In classical (parametric) hypothesis testing many assumptions need to be satisfied regarding distributional properties of the random variables under investigation. Normality of the data is usually assumed when making inferences on location parameters. Parametric and permutation hypothesis tests are investigated in this chapter using the results generated from simulation studies of the power performance of each method. The permutation tests are of particular interest as fewer assumptions are required to be satisfied for proper inference. This is crucial in the multivariate context of multispecies microcosm experiments, particularly since sample sizes are usually very small (10-20) and the data multivariate (often $P \geq N$).

The Hotelling T^2 test on multivariate location vectors is included in the study more as a reference method than as a serious choice as an inferential tool. However, when all assumptions are met, the Hotelling T^2 test is the UMP test for two multivariate location vectors within the class of invariant subspaces (it reduces to the Student's t-test in the univariate case).

The objective of the simulation studies is to shed light on the behavior and performance characteristics of the methodology considered thus far for the analyses of similarity/dissimilarity data. Although we have for the simulations varied a number

of key parameters in an attempt to reflect actual experimental conditions, these studies are in no way exhaustive.

Several simulations studies were conducted. The largest and most comprehensive involved all methods, 1000 replications per simulation and numerous parameter settings. Smaller studies were performed once suitable candidate methods were selected. The following is a list of the simulation studies performed:

(i) General study involving all methods using normal data with 1000 replications per simulation.

(ii) Study investigating effects of rotating the points associated with one treatment group cyclically in the plane around the second set of points. The MRPP and regression methods were highlighted in this study using 1000 replications per simulation.

(iii) Study comparing MRPP and regression methods under the influence of outliers (200 replications per simulation).

(iv) Study comparing MRPP and regression methods using lognormal data with 1000 replications per simulation.

§ 2.3.2 Generating Random Variates

Several types of data can be collected in multispecies experiments to assess the impact of toxic substances and other compounds on the aquatic community. The most prevalent types of data collected are species biomass (or biovolume), abundances and presence-absence. The choice of measurement being determine primarily by the circumstances of the particular experiment. In the sequel only data of the first type will be considered. Species biomass is a measure of the total mass of all organisms on a per species basis. Biomass is a continuous measure which is often modelled to follow a normal distribution. We can represent vectors of P–species biomass as $Y_1, Y_2, \ldots, Y_n$, where $Y_i \in \Re^P$ follows a multivariate normal (MVN) distribution with density given as

$$f(y) = (2\pi)^{-\frac{P}{2}} |\sum|^{-\frac{1}{2}} \exp\left(-\tfrac{1}{2}(y-\mu)'\sum^{-1}(y-\mu)\right) \qquad [2.30]$$

for $-\infty < y_i < \infty$, i=1, 2, ..., P and Σ positive definite, but of course, $y_i \geq 0$ for biomass data. The distribution of Y is completely specified by its mean vector μ and variance–covariance matrix Σ and is usually denoted as $Y \sim MVN_P(\mu, \Sigma)$.

To simulate MVN data we obtained (pseudo) random samples of a specified size from a MVN probability distribution with parameters μ and Σ. This was accomplished by invoking the IMSL FORTRAN subroutine DRNMVN (IMSL 1987).

The subroutine DRNMVN which utilizes double precision arithmetic generates N random P-dimensional vectors in a single invocation, each according to the distribution $MVN_P(\underset{\sim}{0},\Sigma)$. The mean vector μ is then added to each sampled vector so that the resulting data follow $MVN_P(\mu, \Sigma)$.

Prior to the first invocation of DRNMVN, two additional IMSL subroutines were called. The initial seed which controls the data stream of the random variates was set by a call to RNSET. Calling RNSET before any random numbers are generated assures that the resulting stream of numbers will be the same across programs thus allowing for comparisons of methods and different parameter settings under identical conditions. The subroutine DRNMVN requires as input the P×P upper triangular matrix containing the Cholesky square root decomposition of the variance-covariance matrix. That is, it requires as input R, where $S = R'R$ is the variance-covariance matrix for Y.

Table 2.2 details the parameter settings involved in the general simulation study of all the methods. Two independent P-variate normal data vectors of length N_1 and N_2 were generated as described. The N×N symmetric dissimilarity matrix of Euclidean distances was calculated for each set of data ($N = N_1 + N_2$). Entries of this matrix were used as data for all methods except Hotelling's T^2 which required as input the raw values. Statistics were calculated using formulas 2.4, 2.15, 2.17 and 2.21.

One thousand simulations were generated for each method and the number of

Table 2.2. Parameter settings employed in the general simulation study.

Sample Size N_1	Sample Size N_2	# Dependent Variables	Covariance Structure	Location Shift δ
5	5	1	Equal	0, .5, 1, 1.5, 2, 2.5, 3, 4
5	5	1	Unequal	0, .5, 1, 1.5, 2, 2.5, 3, 4
5	5	5	Equal	0, .5, 1, 1.5, 2, 2.5, 3, 4
5	5	5	Unequal	0, .5, 1, 1.5, 2, 2.5, 3, 4
5	5	5	Corr	0, .5, 1, 1.5, 2, 2.5, 3, 4
5	5	20	Equal	0, .5, 1, 1.5, 2, 2.5, 3, 4
5	5	20	Unequal	0, .5, 1, 1.5, 2, 2.5, 3, 4
5	5	20	Corr	0, .5, 1, 1.5, 2, 2.5, 3, 4
5	10	1	Equal	0, .5, 1, 1.5, 2, 2.5, 3, 4
5	10	1	Unequal	0, .5, 1, 1.5, 2, 2.5, 3, 4
5	10	5	Equal	0, .5, 1, 1.5, 2, 2.5, 3, 4
5	10	5	Unequal	0, .5, 1, 1.5, 2, 2.5, 3, 4
5	10	5	Corr	0, .5, 1, 1.5, 2, 2.5, 3, 4
5	10	20	Equal	0, .5, 1, 1.5, 2, 2.5, 3, 4
5	10	20	Unequal	0, .5, 1, 1.5, 2, 2.5, 3, 4
5	10	20	Corr	0, .5, 1, 1.5, 2, 2.5, 3, 4
10	10	1	Equal	0, .5, 1, 1.5, 2, 2.5, 3, 4
10	10	1	Unequal	0, .5, 1, 1.5, 2, 2.5, 3, 4
10	10	5	Equal	0, .5, 1, 1.5, 2, 2.5, 3, 4
10	10	5	Unequal	0, .5, 1, 1.5, 2, 2.5, 3, 4
10	10	5	Corr	0, .5, 1, 1.5, 2, 2.5, 3, 4
10	10	20	Equal	0, .5, 1, 1.5, 2, 2.5, 3, 4
10	10	20	Unequal	0, .5, 1, 1.5, 2, 2.5, 3, 4
10	10	20	Corr	0, .5, 1, 1.5, 2, 2.5, 3, 4

rejections used as an estimate of power. Three covariance structures were utilized in generating the data:

(1) Equal : $\text{Cov}(X) = \text{Cov}(Y) = \sigma_1^2 \cdot I_P$

(2) Unequal : $\text{Cov}(X) = \sigma_1^2 \cdot I_P$ and $\text{Cov}(Y) = \sigma_2^2 \cdot I_P$

(3) Correlation : $\text{Cov}(X) = \text{Cov}(Y) = \sigma_1^2 \cdot I_P + \rho \cdot (J_P - I_P)$,

where X and Y represent N×P matrices of species abundances or biomass, I_P is a P×P identity matrix, J_P is a P×P matrix of ones, $\text{Cov}(X) = \Sigma_X$ and $\text{Cov}(Y) = \Sigma_y$. Values for the variances and correlation coefficient were fixed at $\sigma_1^2 = 1$, $\sigma_2^2 = 3$, and $\rho = 0.8$, respectively. These three covariance structures were used to generate data according to the following distributions:

(i) $X \sim \text{MVN}_P(\mu_X , \Sigma_X)$

(ii) $Y \sim \text{MVN}_P(\mu_y , \Sigma_y)$

where $\mu_X = 0_{1xP}$, $\mu_y = \delta \cdot 1_{1xP}$ for $\delta = 0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 4.0$, and Σ_X and Σ_y are as described previously.

The second simulation study was conducted to investigate variable redundancy. The set of points associated with one sample was rotated cyclically in the plane, at a fixed distance, around the set of points for the second sample as depicted in Figure 2.1. Two independent and identically distributed sets of bivariate normal observation vectors X and Y of size 5 were generated as described in the previous paragraph. To each element of the first and second dimension of Y was added $\delta \cdot \text{COS}(X \cdot \frac{\pi}{8})$ and $\delta \cdot \text{SIN}(X \cdot \frac{\pi}{8})$, respectively, for $\delta = 1, 2, 3$ and $X = 1, \ldots, 16$ so that the set of points for Y when plotted in the plane would cyclically rotate around the set of points for X in $\frac{\pi}{8}$

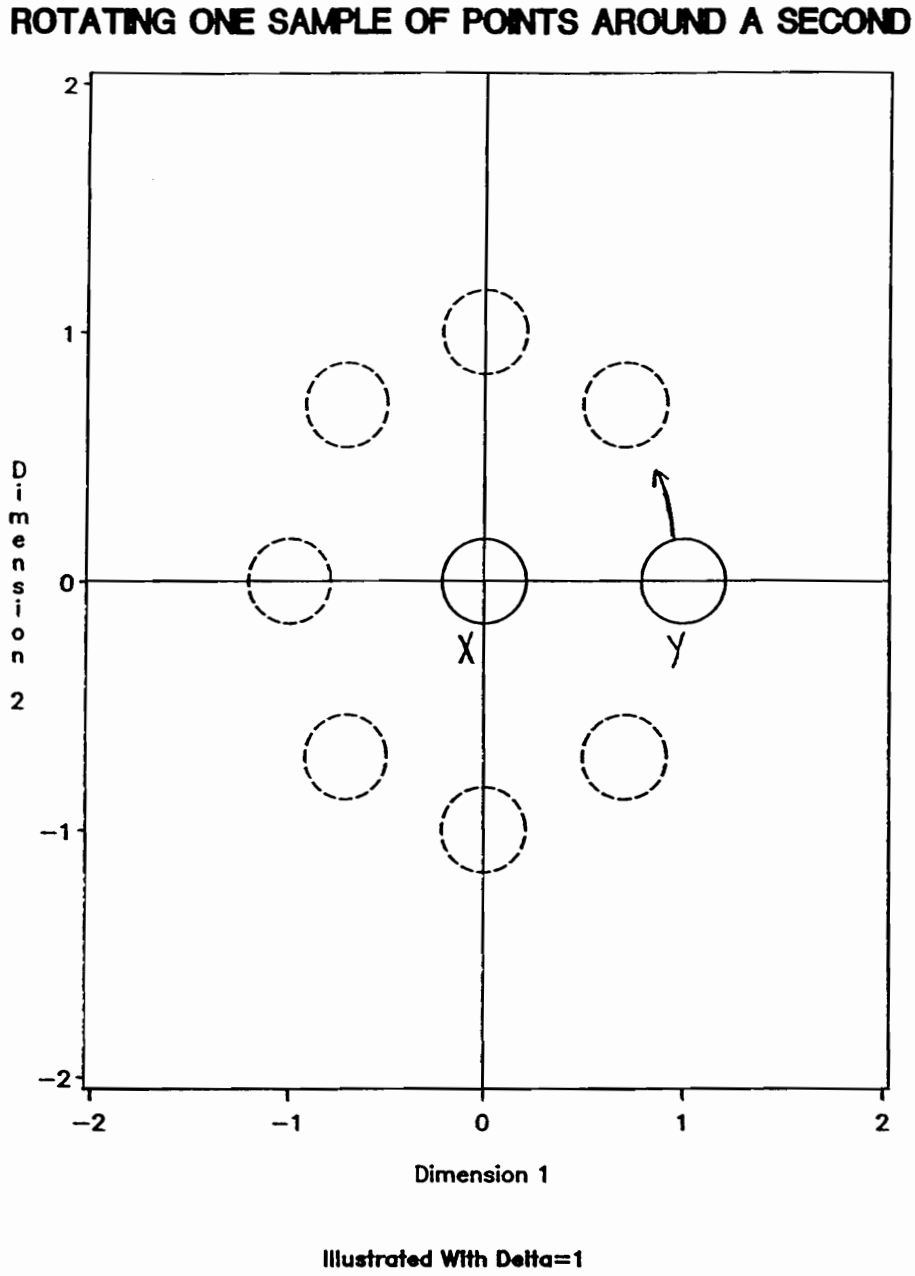

Figure 2.1. Data points from one sample rotated cyclically in the plane around points from the second sample.

increments (δ provided the necessary separation in the centroids of X and Y). It was anticipated that there would be certain orientations of the data that would essentially make the information in one of the sets of points redundant thereby decreasing test performance.

MRPP and regression methods were used to analyze the data after summarization into dissimilarity matrices. All three covariance structures were used in generating data for the 1000 simulations performed.

The third study involved the MRPP and regression methods in the investigation of the effect of an outlying observation on test performance. Two independent and identically distributed sets of bivariate normal observation vectors X and Y of size 10 were generated under all three covariance structures as described previously (200 simulations). All points in Y were translated along the abscissa by $\delta = 1$ or 2 units to provide separation in the centroids of X and Y. To create the outlier, a single observation from Y was further perturbed by adding the point $(\zeta,0)$, $(0,\zeta)$, $(-\zeta,0)$ or $(0,-\zeta)$, for $\zeta=2$, 4 (Figure 2.2). This effectively rotated the point through each quadrant when viewing the original point as the origin and allowed for the effects of this point on test performance of the MRPP and regression methods to be studied.

The final study involving generated data was designed to compare the MRPP and regression methods under a plausible, non-normal underlying distribution. Two samples of independent P–dimensional lognormal random observation vectors X and Y of length 5 were obtained for each of 1000 simulations by invoking the double

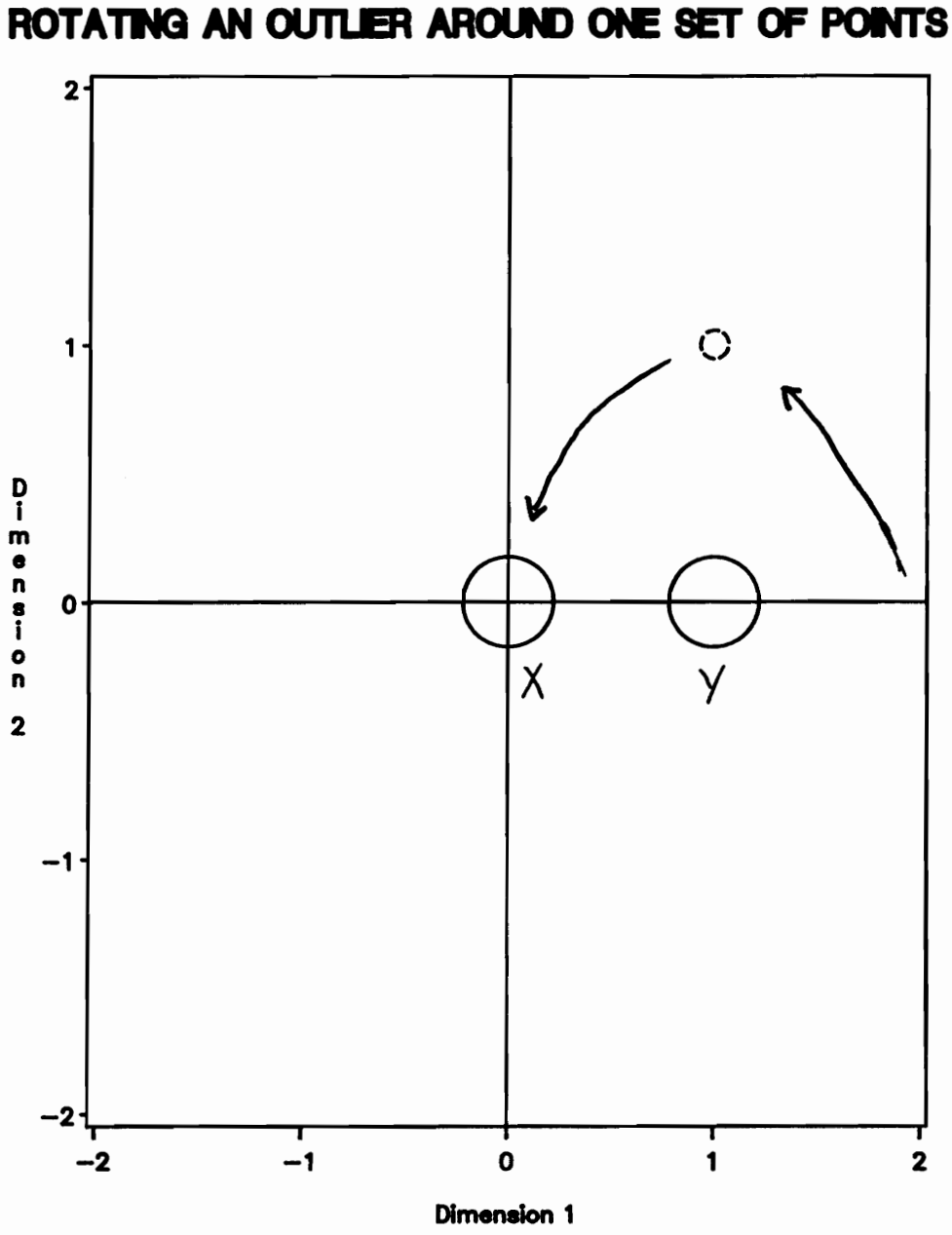

Figure 2.2. Graphical representation of locations of outliers in relation to clusters of observations in X and Y.

precision IMSL subroutine DRNLNL to generate univariate lognormal variates distributed as $X \sim LN(\exp(0.5), e(e-1))$ and $Y \sim LN(\exp(\mu+\sigma^2/2), \exp(2\mu+2\sigma^2) - \exp(2\mu+\sigma^2))$, where the underlying distributions are for $X \sim N(0,1)$ and $Y \sim N(\mu,\sigma^2)$; $P = 1, 5$, $\mu = 0, 0.5, \ldots, 3.0$ and $\sigma^2 = 1, 3$.

§ 2.3.3 Results

§ 2.3.3.*A* *General Simulation—Test Validity: α-level Concerns*

Assessment of test validity was determined by examination of the α-levels for all procedures under the equal covariance structure. In general, α-levels held very close to the *a priori* selected error rate of $\alpha = 0.05$ and seemed to be unaffected by the presence of correlation when the covariance structures were identical in both groups.

Significant departures from $\alpha = 0.05$ were observed under the unequal covariance structure (Figure 2.3). The most severe α-level inflation occurred when $P > 1$, $N_1 \neq N_2$ (see Figures 2.15 and 2.18). Under these conditions, α-levels averaged 50% for the regression method and 75% for the ANOVA method while the MRPP and Hotelling T^2 procedures held the α-levels to 0.05 under all conditions (Table 2.3). In one sense, the inflation of α-levels under unequal covariance structures can be viewed as a desirable characteristic. Unequal covariance structures usually infers different underlying distributions for the two groups, and it would seem desirable for tests to

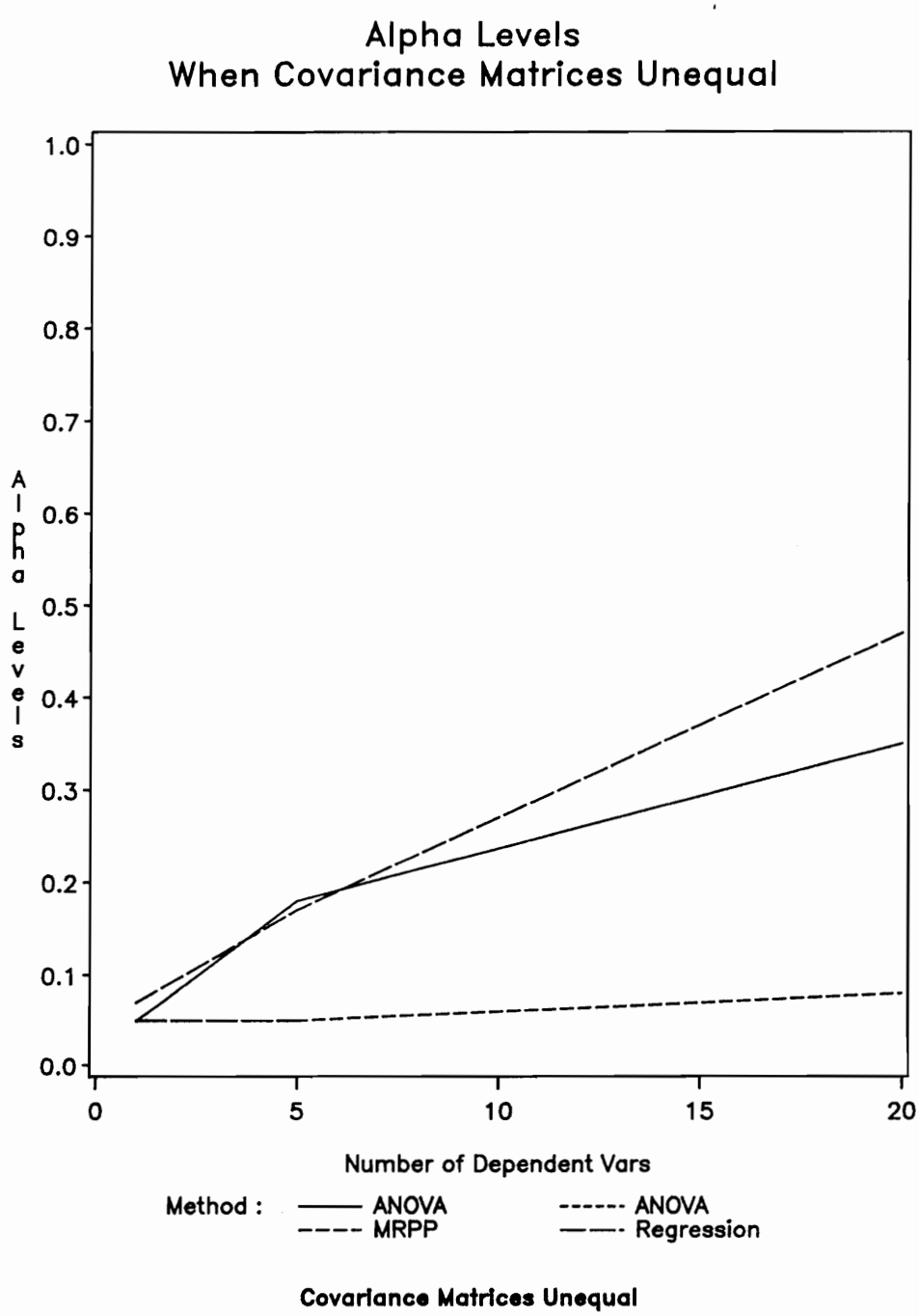

Figure 2.3. Observed α—levels for the general simulation when the covariance matrices are unequal. Note that Hotelling T^2 cannot be computed for P=20.

Table 2.3. Observed α-levels from the general simulation study.

Sample Size N_1	Sample Size N_2	# Dependent Variables	Covariance Structure	Tests[1] MRPP	Regr	AOV	Hotelling
5	5	1	Equal	5	5	5	5
5	5	1	Unequal	5	5	5	5
5	5	5	Equal	5	5	5	5
5	5	5	Unequal	5	10	5	5
5	5	5	Corr	5	5	5	5
5	5	20	Equal	5	5	5	–
5	5	20	Unequal	5	5	5	–
5	5	20	Corr	5	5	5	–
5	10	1	Equal	5	5	5	5
5	10	1	Unequal	5	5	5	5
5	10	5	Equal	5	5	5	5
5	10	5	Unequal	5	20	45	5
5	10	5	Corr	5	5	5	5
5	10	20	Equal	5	5	10	–
5	10	20	Unequal	5	80	95	–
5	10	20	Corr	5	5	5	–
10	10	1	Equal	5	5	5	5
10	10	1	Unequal	5	10	5	5
10	10	5	Equal	5	5	5	5
10	10	5	Unequal	5	20	5	5
10	10	5	Corr	5	5	5	5
10	10	20	Equal	5	5	5	–
10	10	20	Unequal	15	55	5	–
10	10	20	Corr	5	5	5	–

[1] α-levels on a percent basis rounded to multiples of 5%.

have the ability to detect this condition.

Immediately following is a discussion of results on the performance of each method under the 24 combinations of sample size, number of dependent variables and covariance structure listed in Table 2.2 for the large, general simulation study. Plots for each of the 24 parameter combinations appear following the discussion (Figures 2.4–2.27). Discussions and supporting plots for the 3 smaller simulation studies follow in order.

§ 2.3.3.*B* *General Simulation—Method Comparisons*

(i) MRPP : The MRPP demonstrated consistently strong power performance under the variety of sample sizes and covariance structures considered. It exhibited power curves very similar to those of the regression method in the $N_1 = N_2$ case, and was not as adversely affected as the regression method for the $N_1 \neq N_2$ case. Degradation in power of the MRPP was slight when $N_1 \neq N_2$ and $Cov(X) \neq Cov(Y)$ for values of $1 \leq \delta \leq 3$. Its performance appears to be superior to the other methods considered under these circumstances.

An unappealling aspect of the MRPP, as well as any permutation–based procedure, is that no account of the *magnitude* of treatment or location differences is made. Whether a treatment difference is minute or huge is in no way reflected by the procedure and not recoverable information.

(ii) Regression Method : Under equal sample sizes the regression method performed as well as the MRPP with respect to power and appeared to be quite robust against the unequal covariance structure. As anticipated it performed quite well when correlation among the dependent variables was present since this was the only procedure which actually models correlation.

Severe deterioration of power was observed when $N_1 \neq N_2$, and was worse when $N_1 \neq N_2$ and $Cov(X) \neq Cov(Y)$, averaging a 30% decrease in power from the equal N case for $1 \leq \delta \leq 3$. In this scenario, α-levels were inflated to an average of $\bar{\alpha} = 34\%$ while the power curve exhibited erratic behavior. Unlike the MRPP, the regression method does account for the magnitude of treatment differences which is reflected in the regression coefficients.

(iii) Hotelling T^2 : This method displayed the best power characteristics in the univariate case (P=1) and the worst when $P \geq 5$. As with the other procedures, power increased positively with increasing sample size and number of dependent variables.

One significant drawback of this procedure is that it can only be used when $N \geq P+2$. In this first simulation, Hotelling's T^2 could not be used for the cases when P=20 (about $\frac{1}{3}$ of the time) as N was always less than P+2. The utility of the T^2 tests is thus severely limited for uses in multispecies microcosm experiments as typically the number of dependent variables is large. Additionally, unequal covariances for the two groups appeared to damage the performance of

this test more severely than the others.

(iv) ANOVA Procedure : Power of the ANOVA procedure often fell considerably below the power achieved by the leading methods, MRPP and regression. As was true for the regression method, it was quite sensitive to imbalance and exhibited significant degradation in power when $N_1 \neq N_2$ and $Cov(X) \neq Cov(Y)$. This would be expected as ANOVA is usually considered a special case of regression. It did, however, perform better than the Hotelling T^2 when $P \geq 5$.

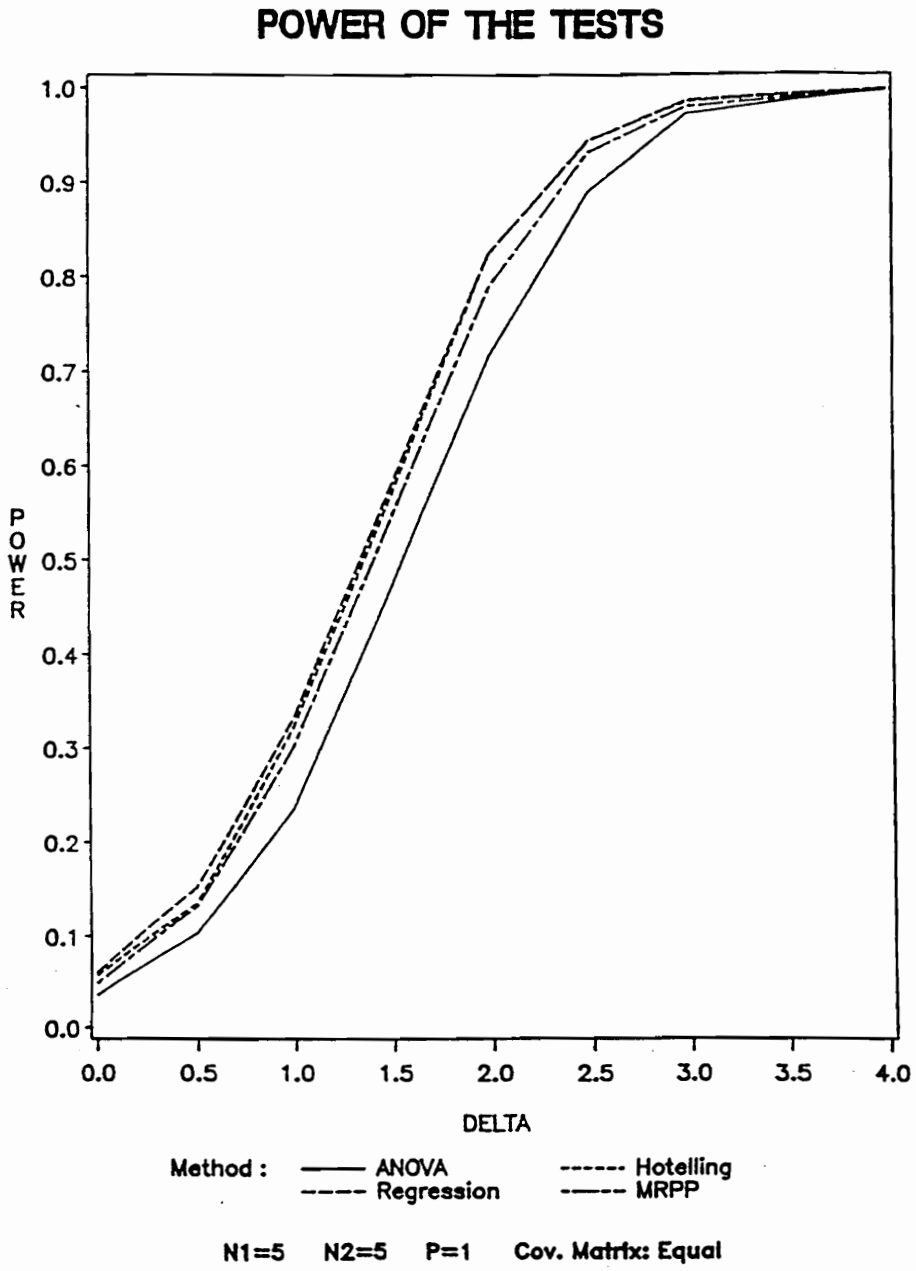

Figure 2.4. Power of the two-sample ANOVA, Regression, Hotelling T^2 and MRPP methods with no correlation present, covariance matrices equal, sample sizes of $N_1=5$ $N_2=5$ and $P=1$.

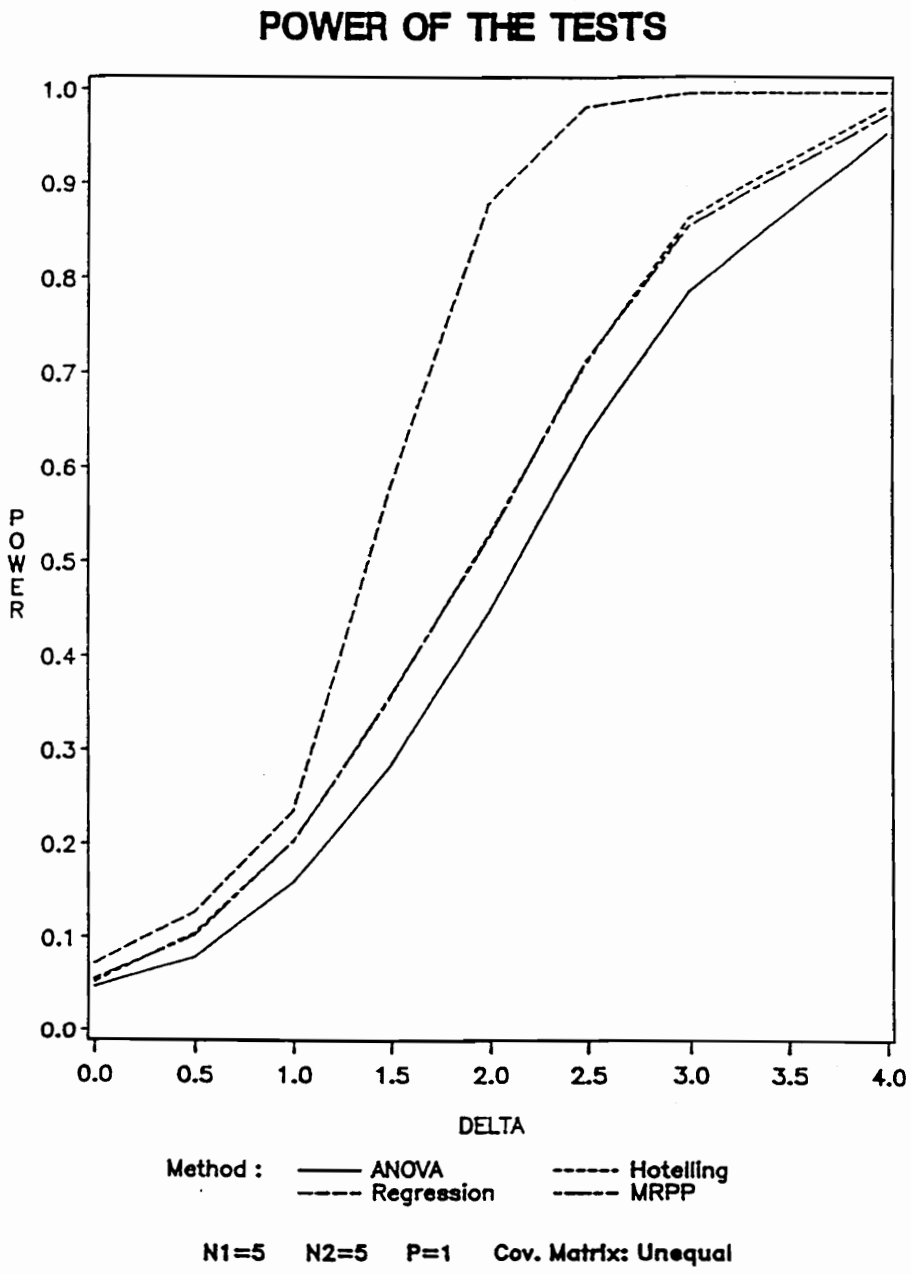

Figure 2.5. Power of the two-sample ANOVA, Regression, Hotelling T^2 and MRPP methods with no correlation present, covariance matrices unequal, sample sizes of $N_1=5$ $N_2=5$ and $P=1$.

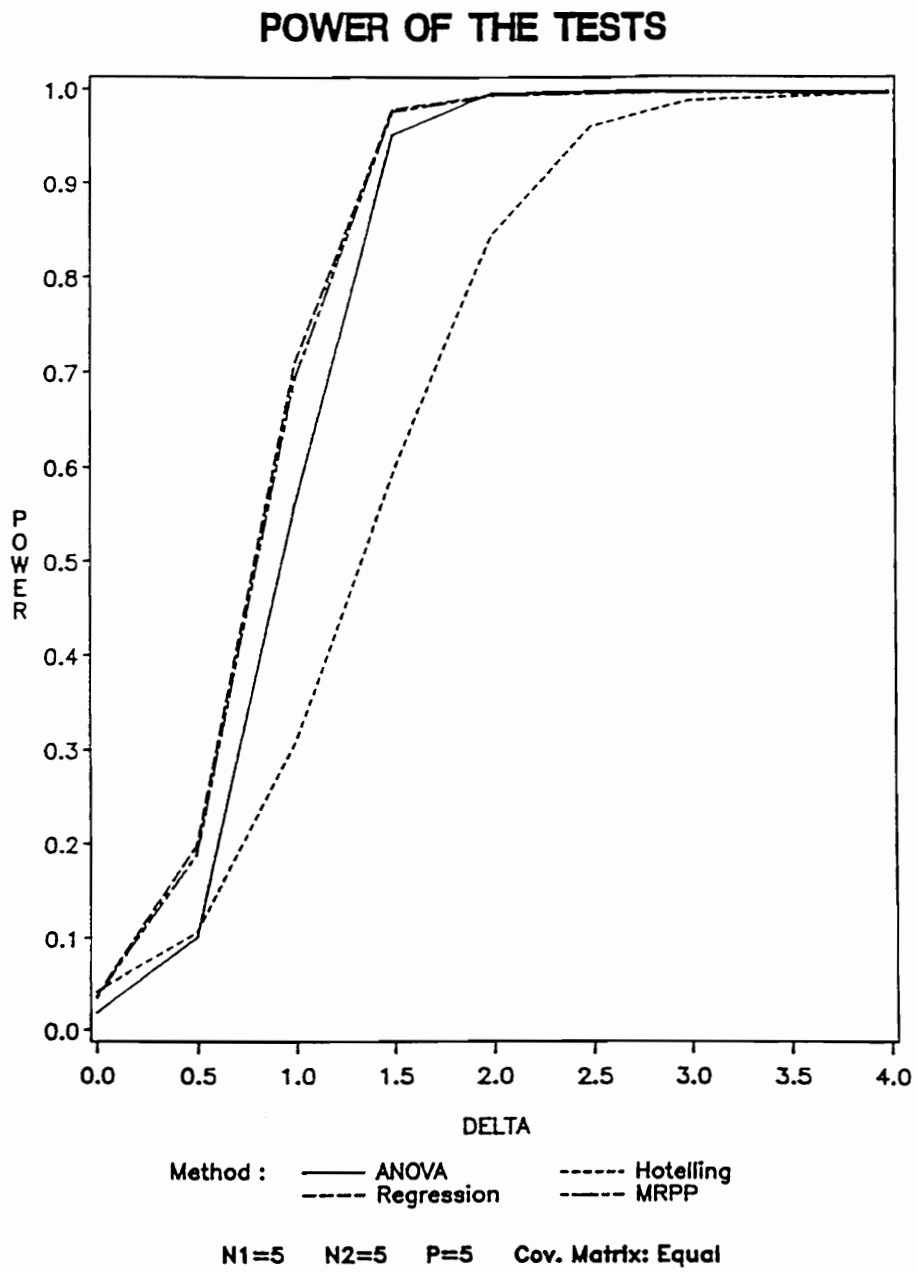

Figure 2.6. Power of the two-sample ANOVA, Regression, Hotelling T^2 and MRPP methods with no correlation present, covariance matrices equal, sample sizes of $N_1=5$ $N_2=5$ and $P=5$.

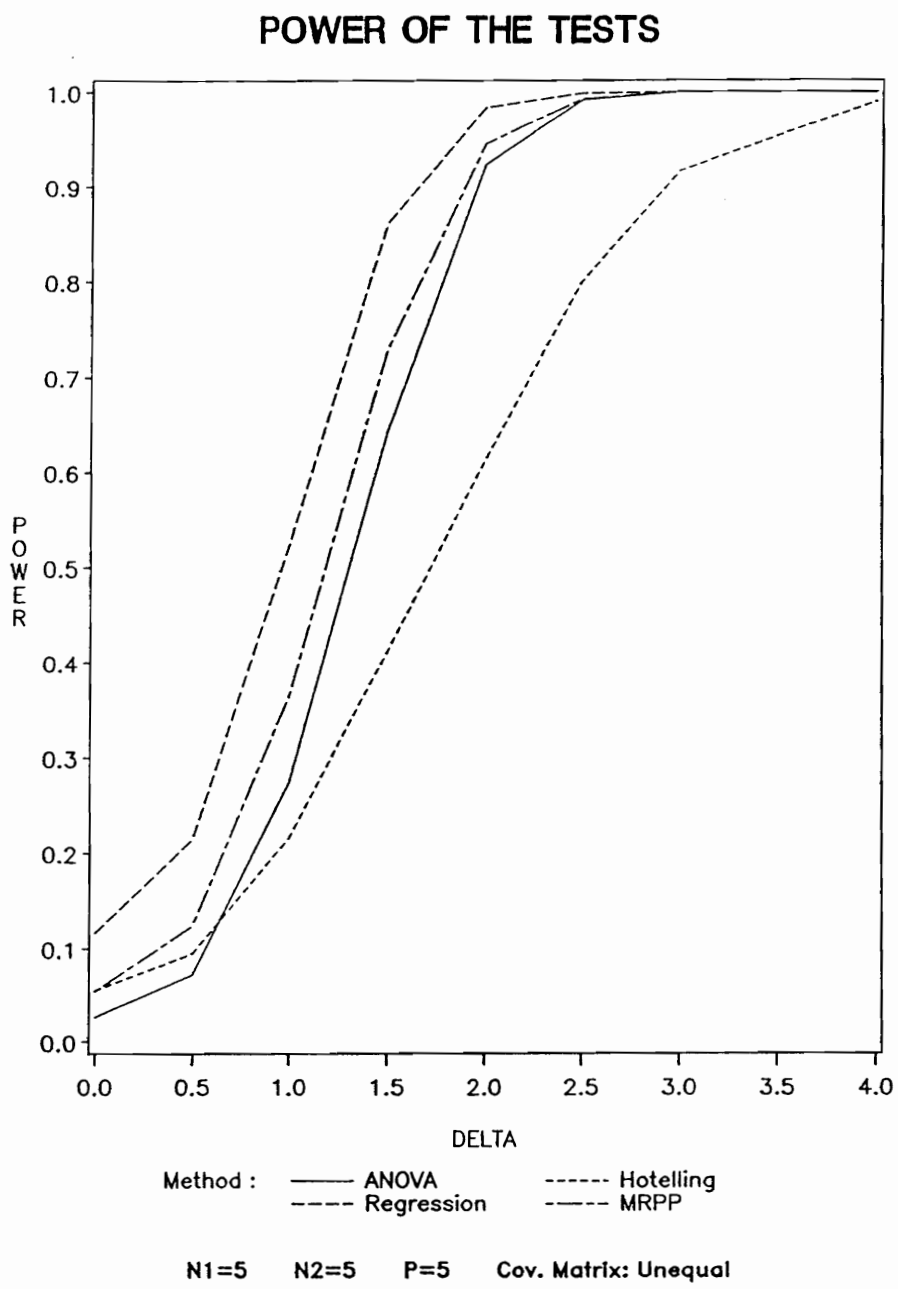

Figure 2.7. Power of the two-sample ANOVA, Regression, Hotelling T^2 and MRPP methods with no correlation present, covariance matrices unequal, sample sizes of $N_1=5$ $N_2=5$ and $P=5$.

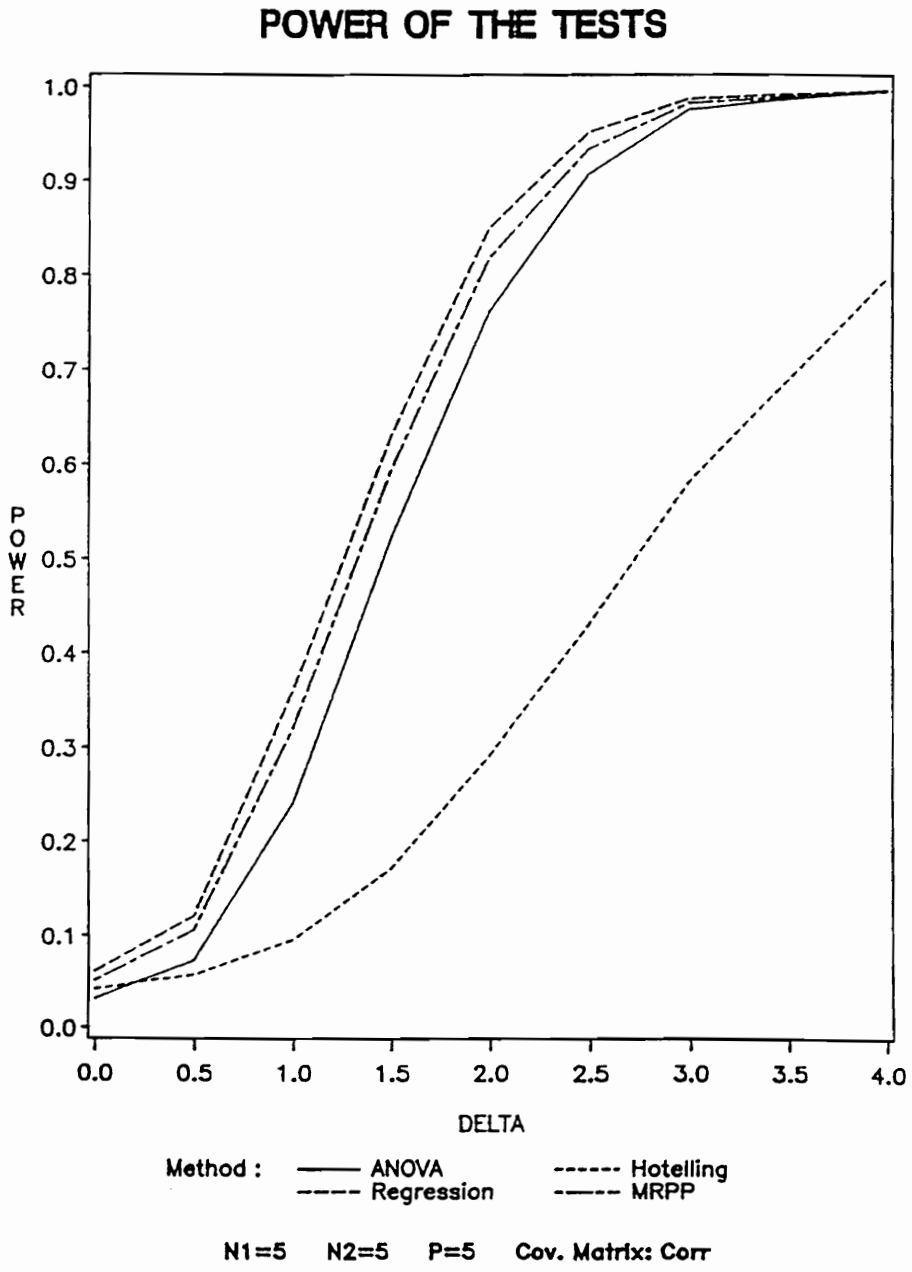

Figure 2.8. Power of the two-sample ANOVA, Regression, Hotelling T^2 and MRPP methods with correlation present (ρ=0.8), covariance matrices equal, sample sizes of N_1=5 N_2=5 and P=5.

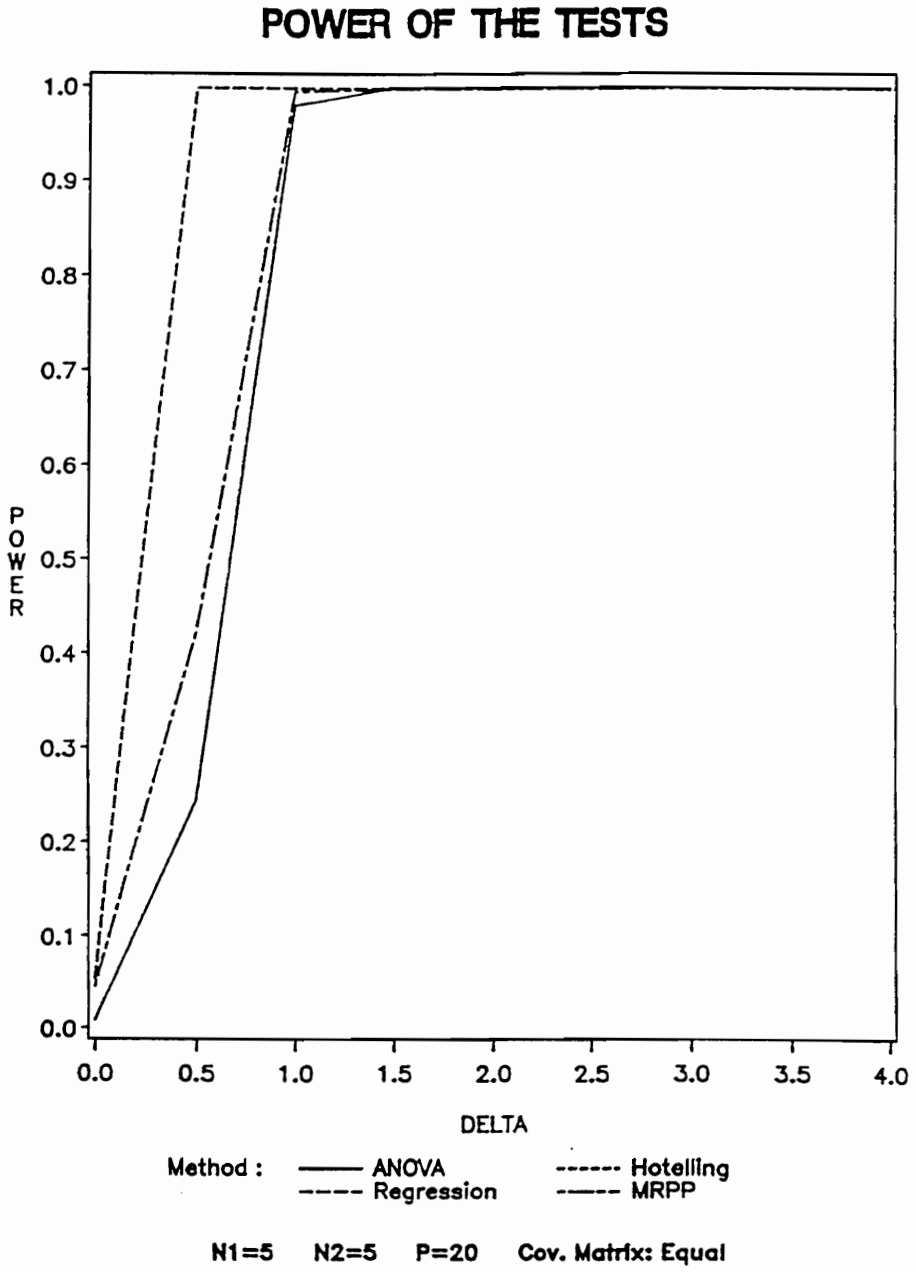

Figure 2.9. Power of the two-sample ANOVA, Regression and MRPP methods with no correlation present, covariance matrices equal, sample sizes of $N_1=5$ $N_2=5$ and $P=20$.

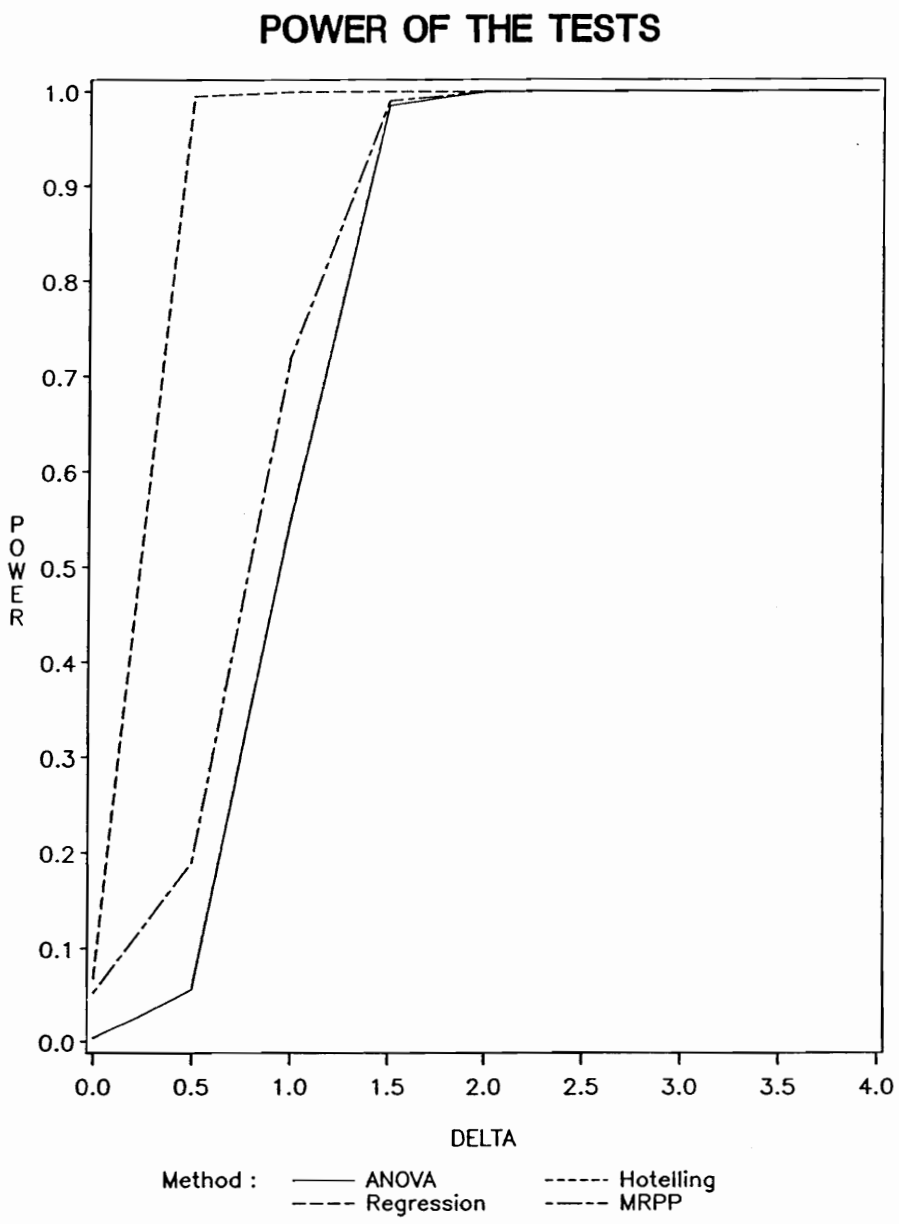

Figure 2.10. Power of the two-sample ANOVA, Regression and MRPP methods with no correlation present, covariance matrices unequal, sample sizes of $N_1=5$ $N_2=5$ and $P=20$.

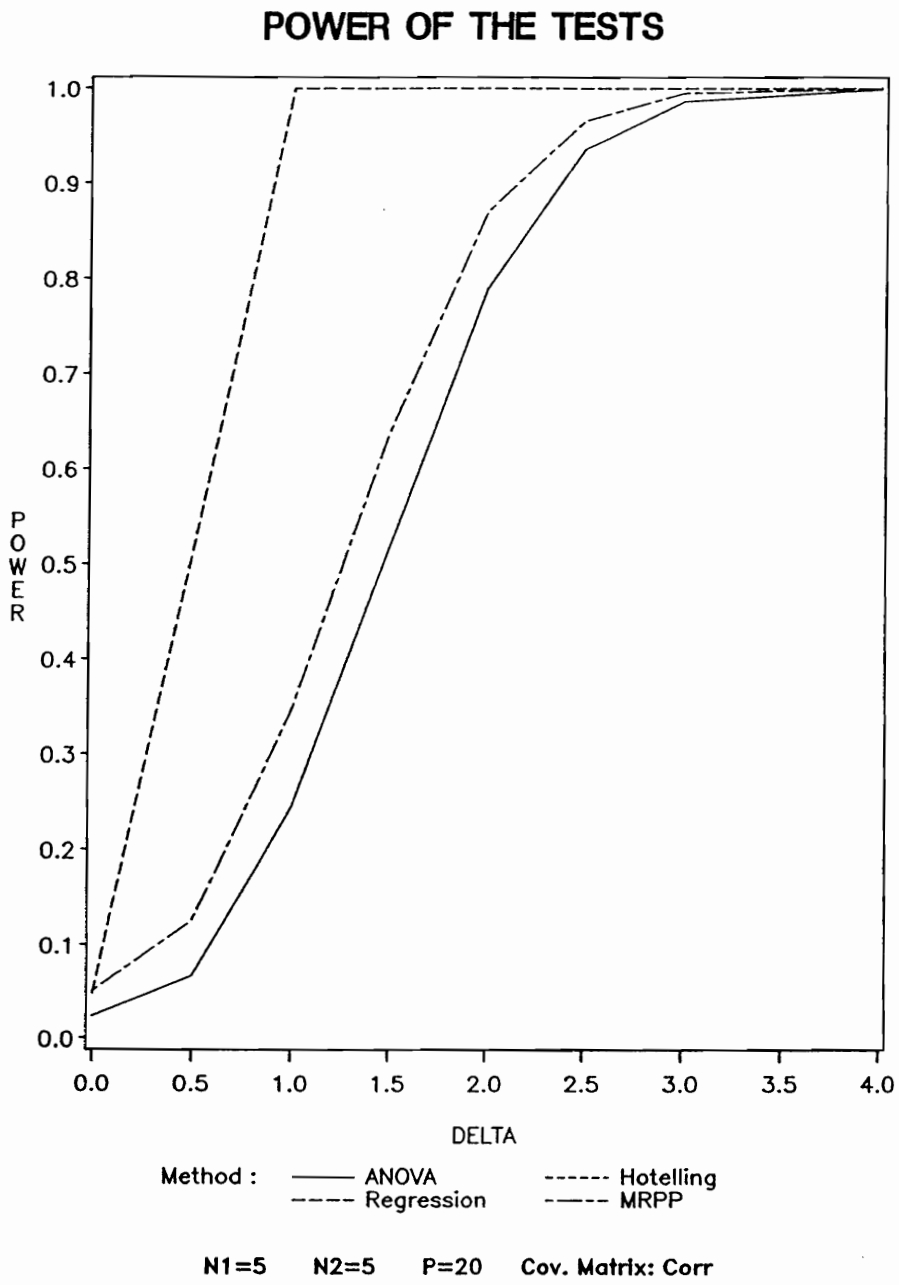

Figure 2.11. Power of the two-sample ANOVA, Regression and MRPP methods with correlation present (ρ=0.8), covariance matrices equal, sample sizes of N_1=5 N_2=5 and P=20.

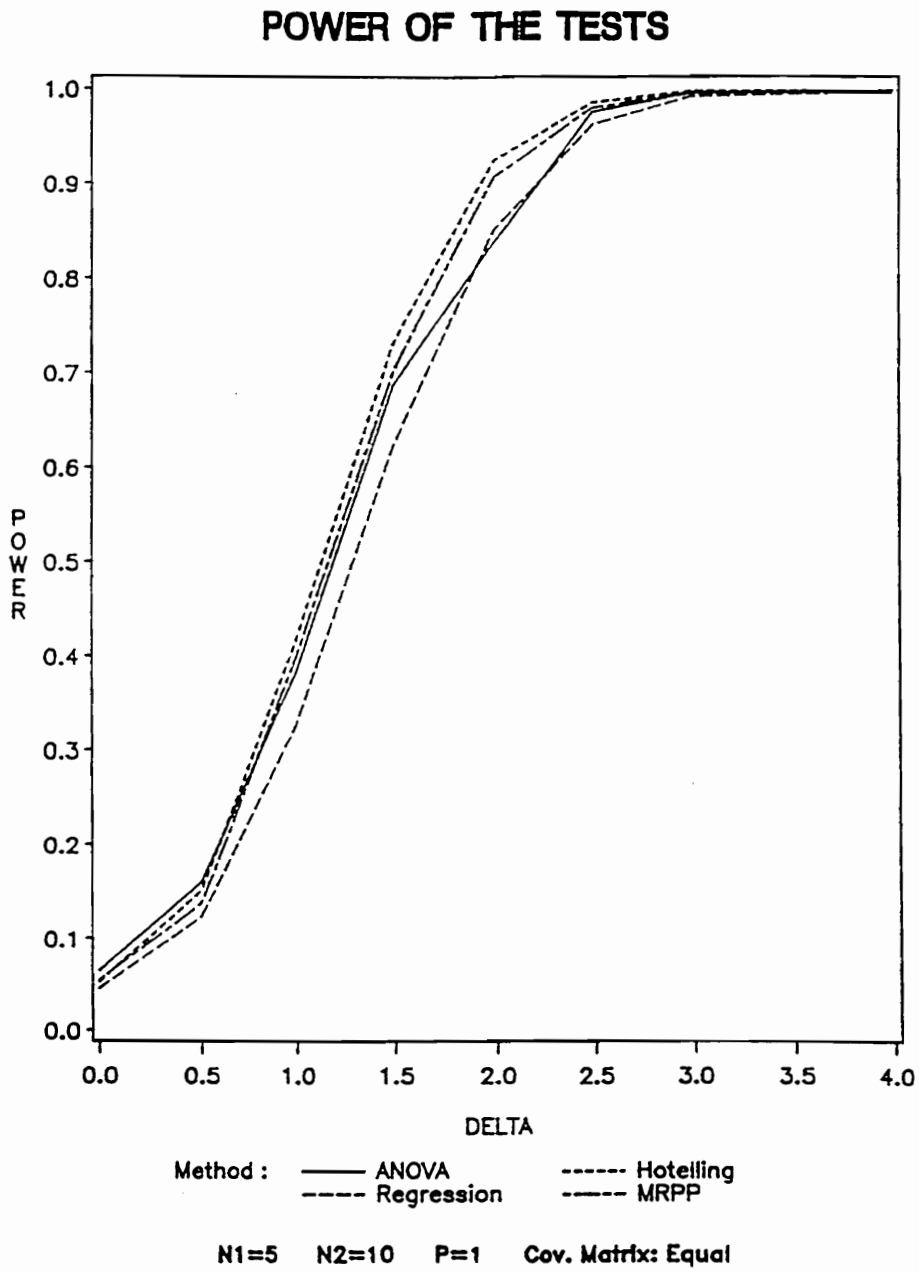

Figure 2.12. Power of the two-sample ANOVA, Regression, Hotelling T^2 and MRPP methods with no correlation present, covariance matrices equal, sample sizes of $N_1=5$ $N_2=10$ and $P=1$.

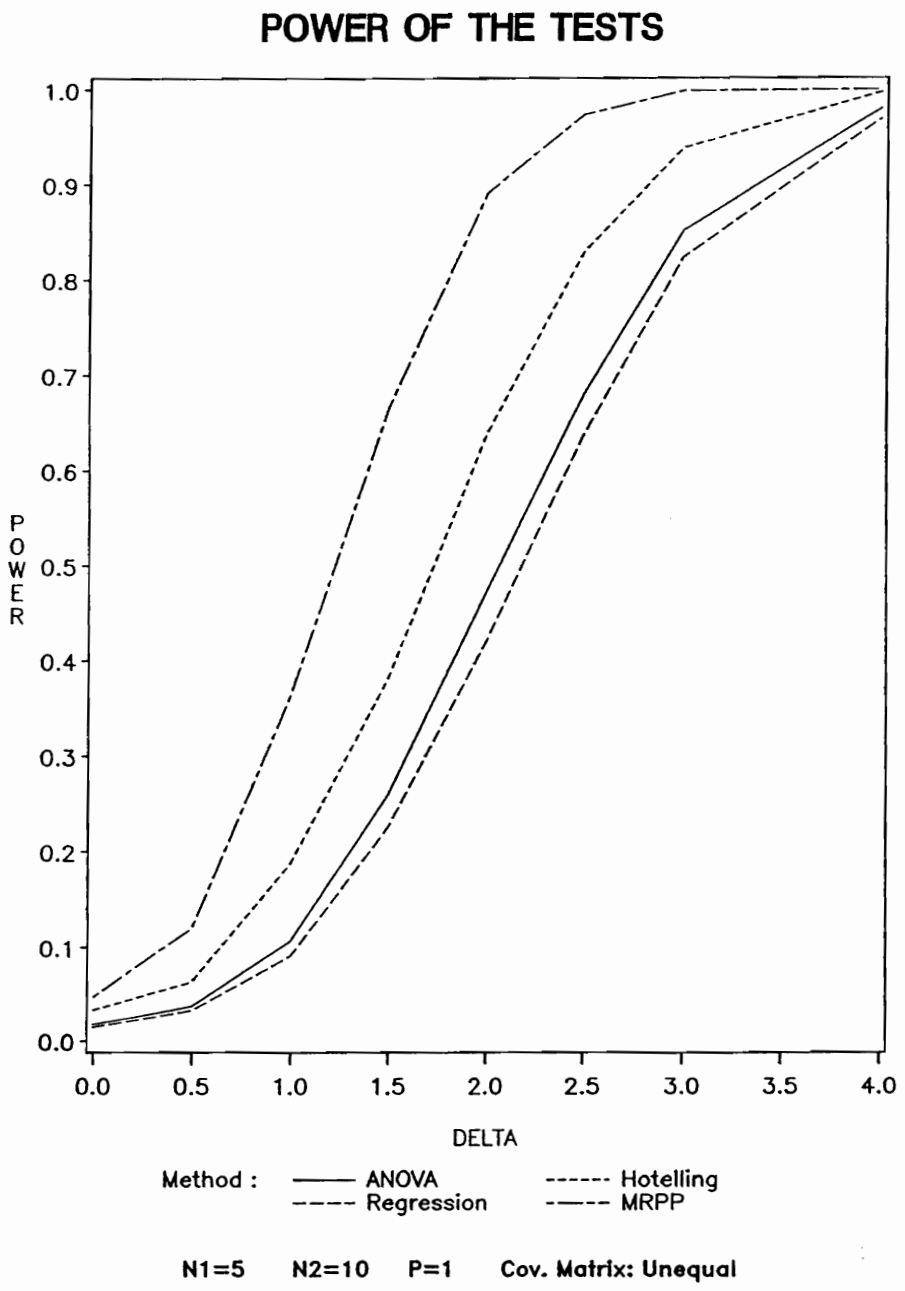

Figure 2.13. Power of the two-sample ANOVA, Regression, Hotelling T^2 and MRPP methods with no correlation present, covariance matrices unequal, sample sizes of $N_1=5$ $N_2=10$ and $P=1$.

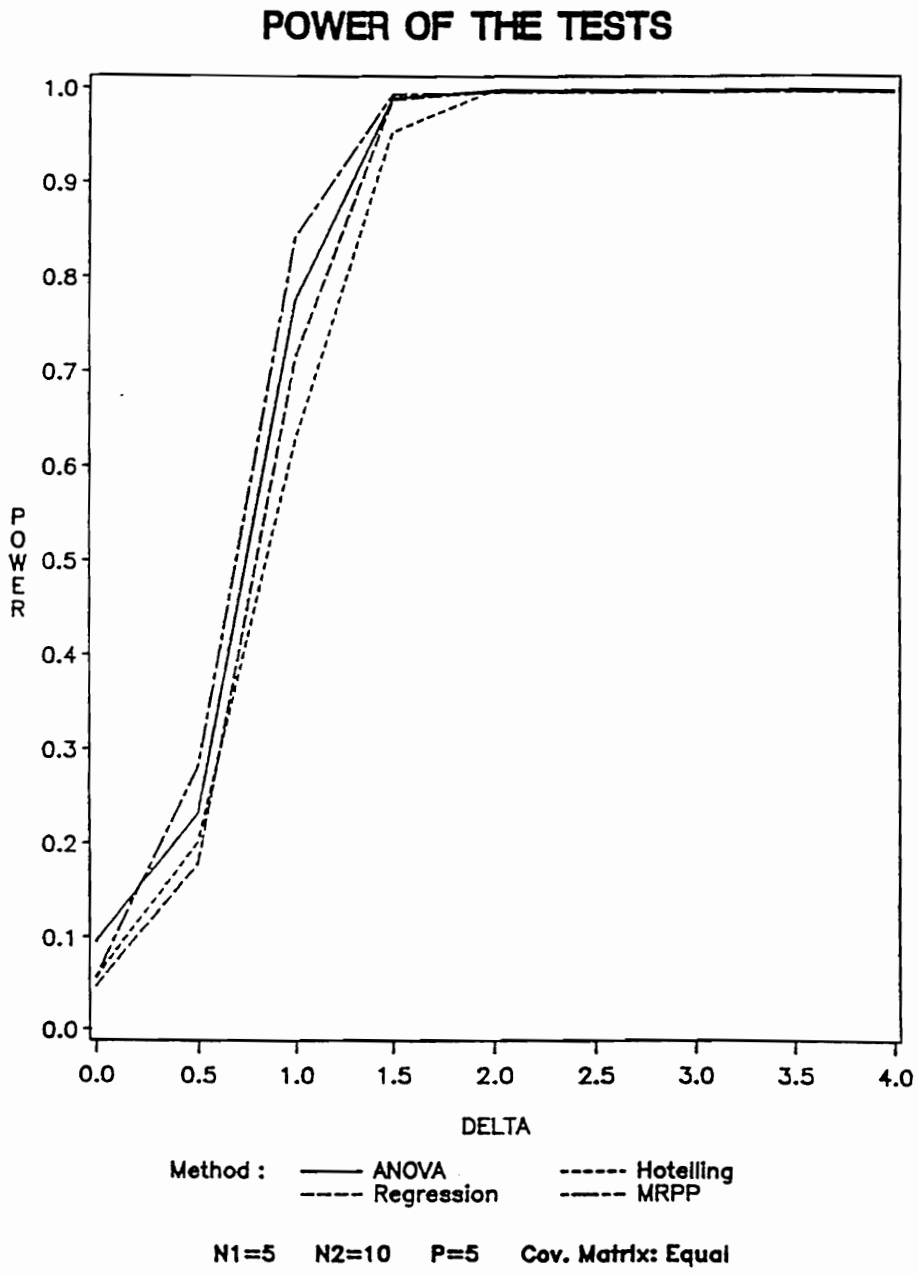

Figure 2.14. Power of the two-sample ANOVA, Regression, Hotelling T^2 and MRPP methods with no correlation present, covariance matrices equal, sample sizes of $N_1=5$ $N_2=10$ and $P=5$.

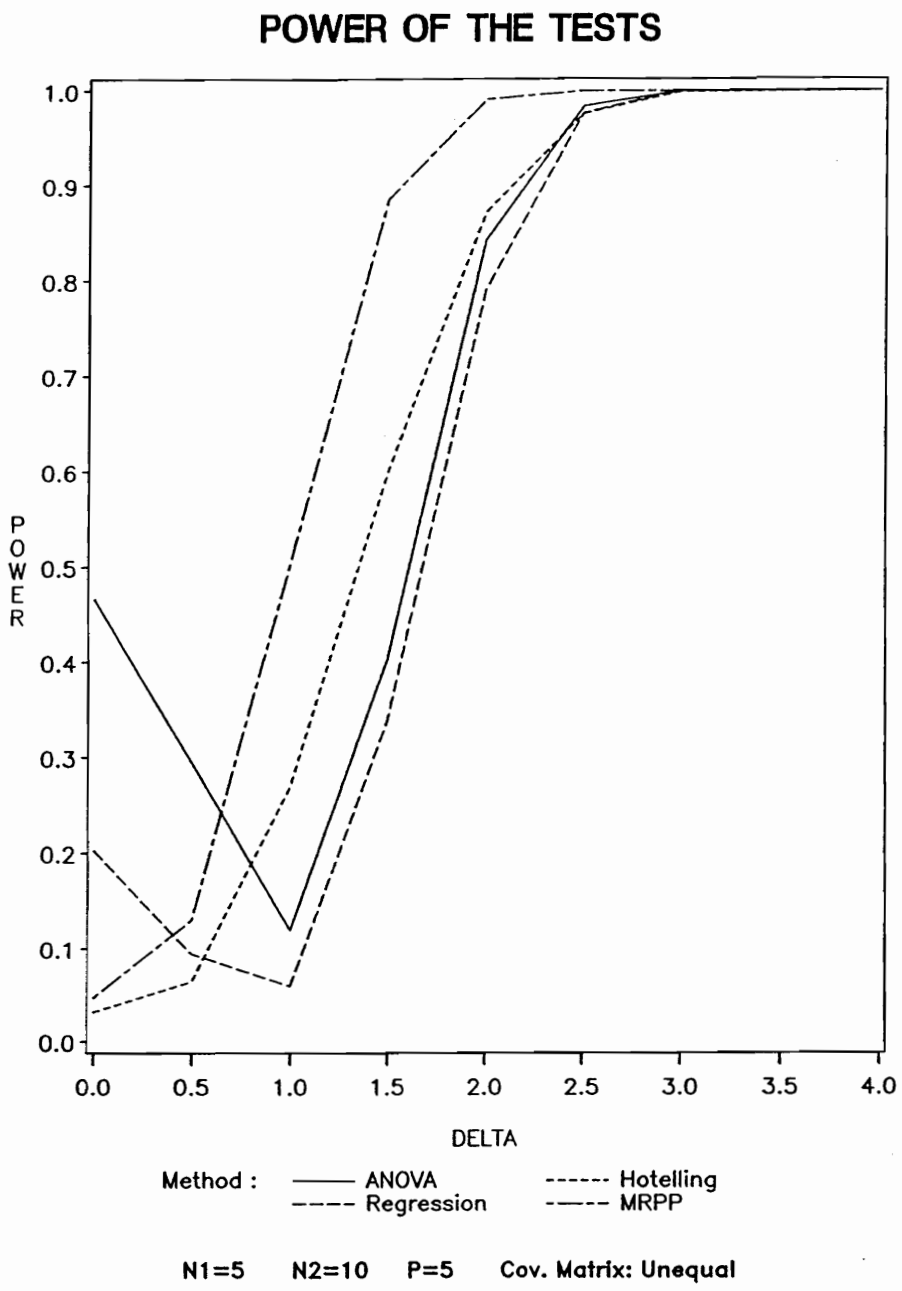

Figure 2.15. Power of the two-sample ANOVA, Regression, Hotelling T^2 and MRPP methods with no correlation present, covariance matrices unequal, sample sizes of $N_1=5$ $N_2=10$ and $P=5$.

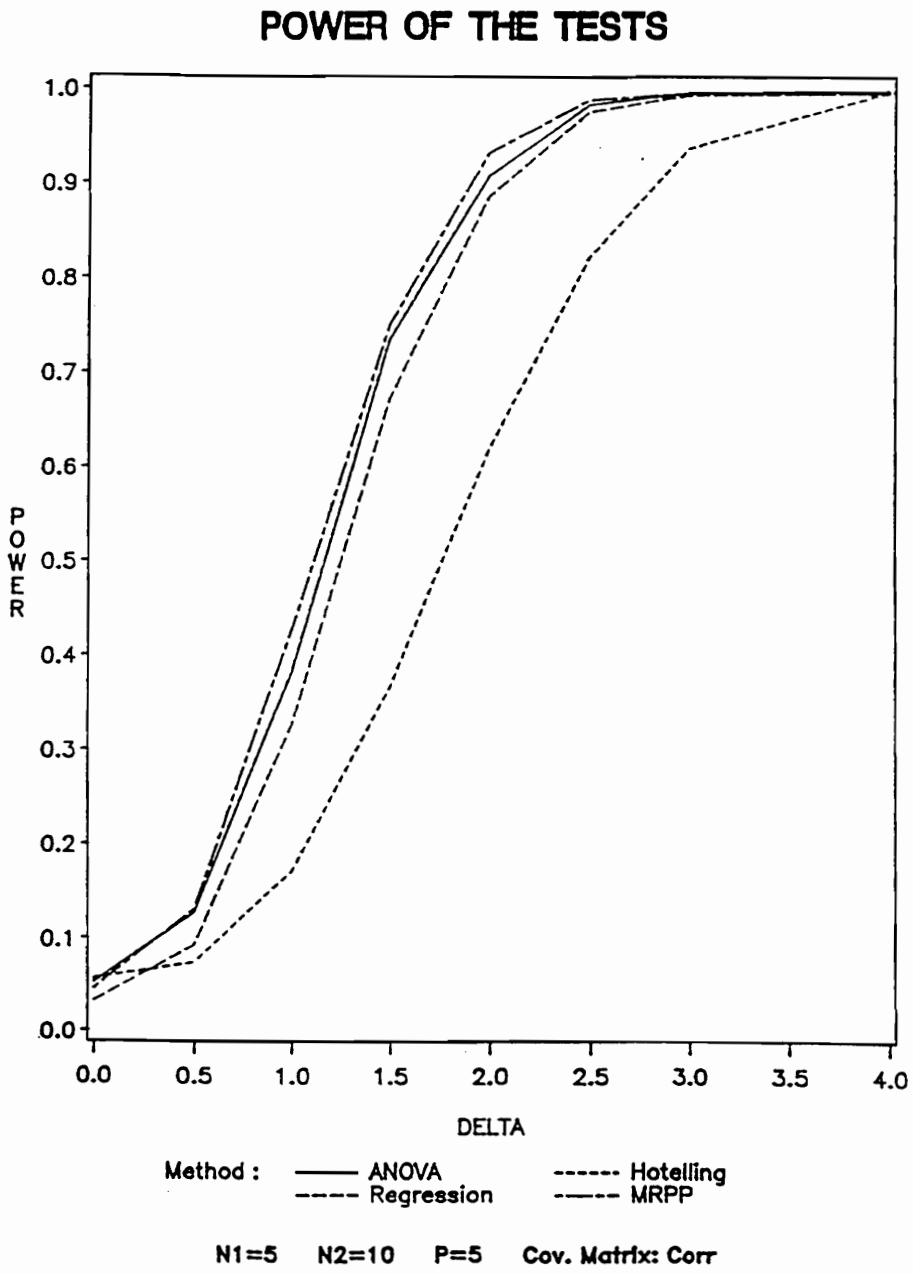

Figure 2.16. Power of the two-sample ANOVA, Regression, Hotelling T^2 and MRPP methods with correlation present (ρ=0.8), covariance matrices equal, sample sizes of N_1=5 N_2=10 and P=5.

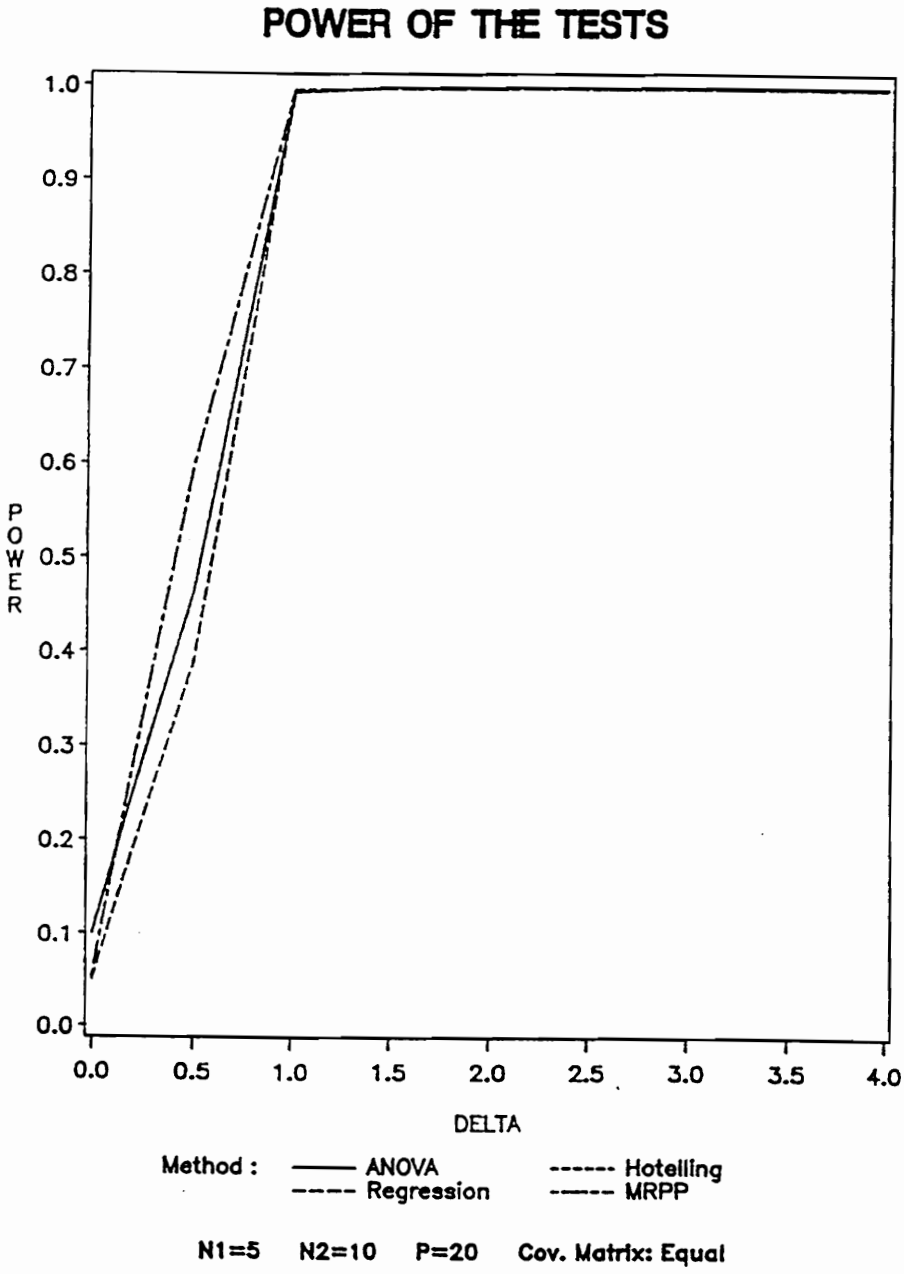

Figure 2.17. Power of the two-sample ANOVA, Regression and MRPP methods with no correlation present, covariance matrices equal, sample sizes of $N_1=5$ $N_2=10$ and $P=20$.

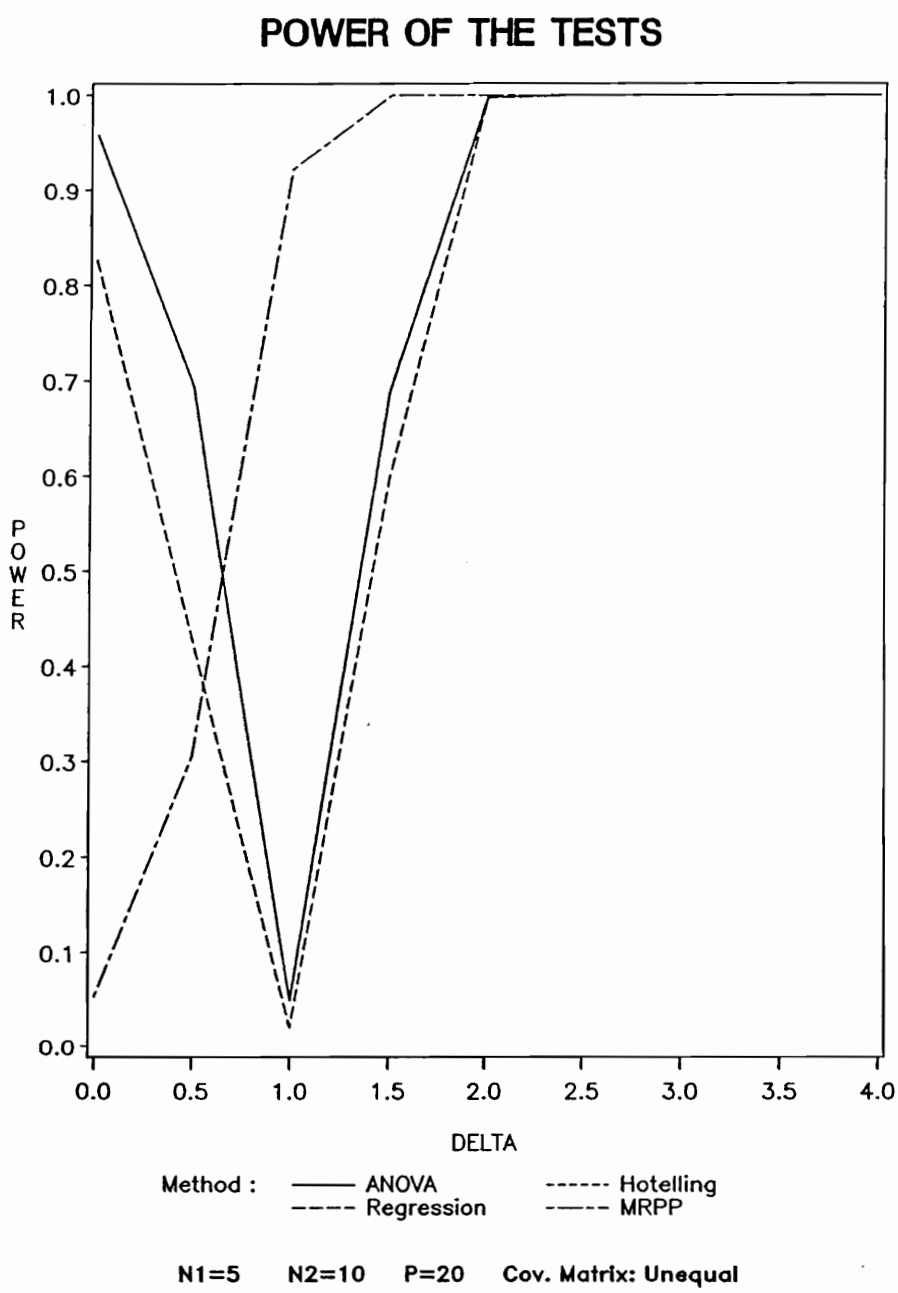

Figure 2.18. Power of the two-sample ANOVA, Regression and MRPP methods with no correlation present, covariance matrices unequal, sample sizes of N_1=5 N_2=10 and P=20.

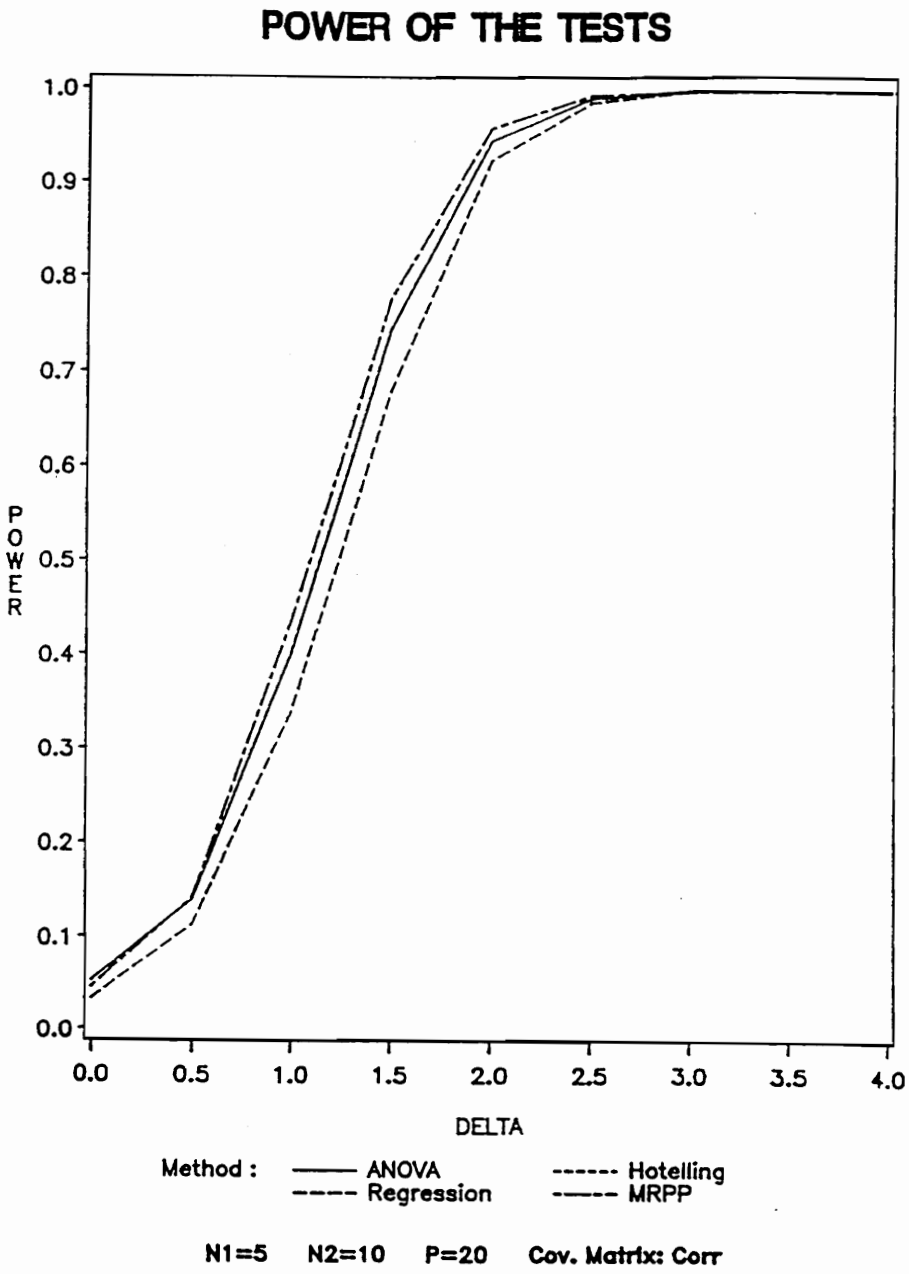

Figure 2.19. Power of the two-sample ANOVA, Regression and MRPP methods with correlation present (ρ=0.8), covariance matrices equal, sample sizes of N_1=5 N_2=10 and P=20.

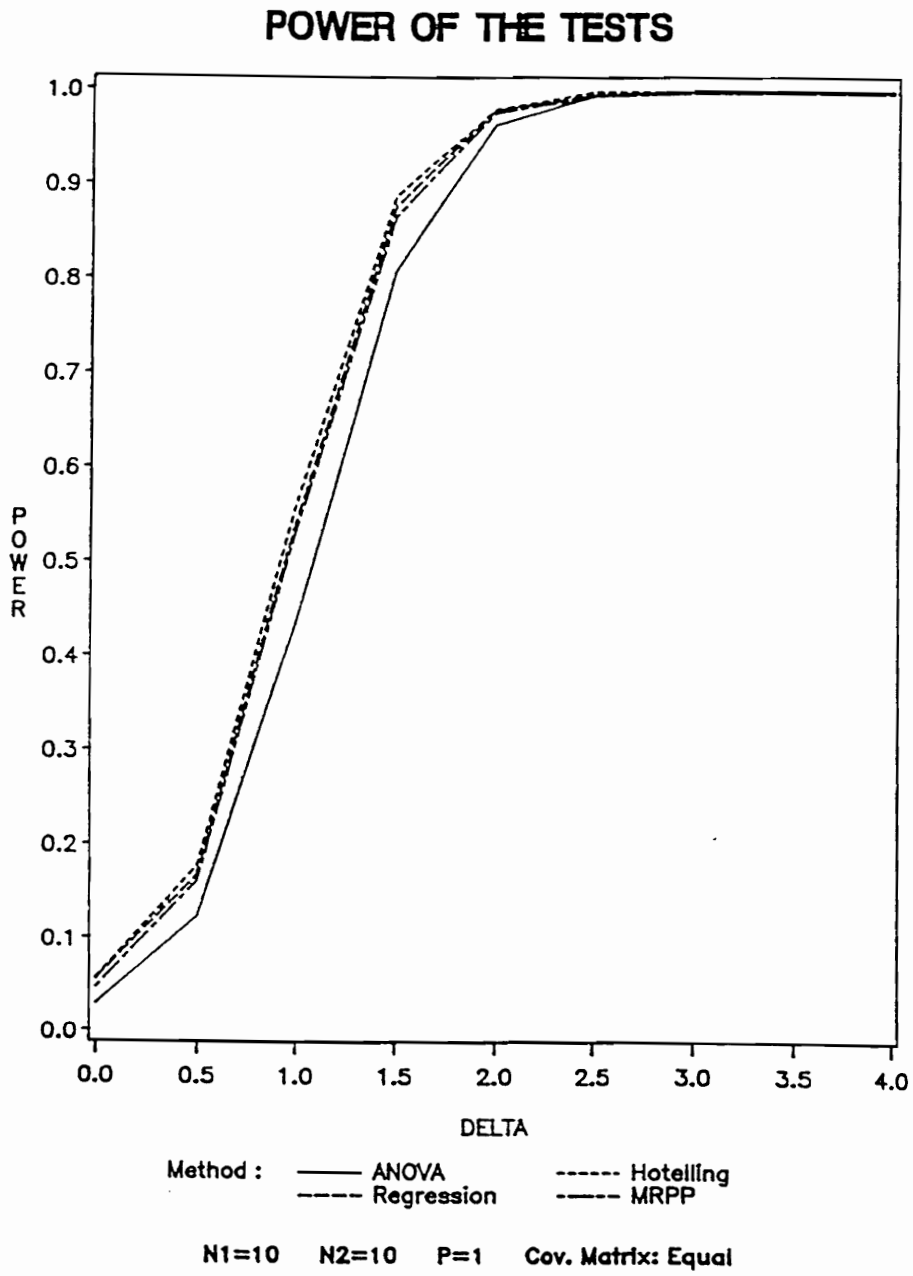

Figure 2.20. Power of the two-sample ANOVA, Regression, Hotelling T^2 and MRPP methods with no correlation present, covariance matrices equal, sample sizes of $N_1=10$ $N_2=10$ and $P=1$.

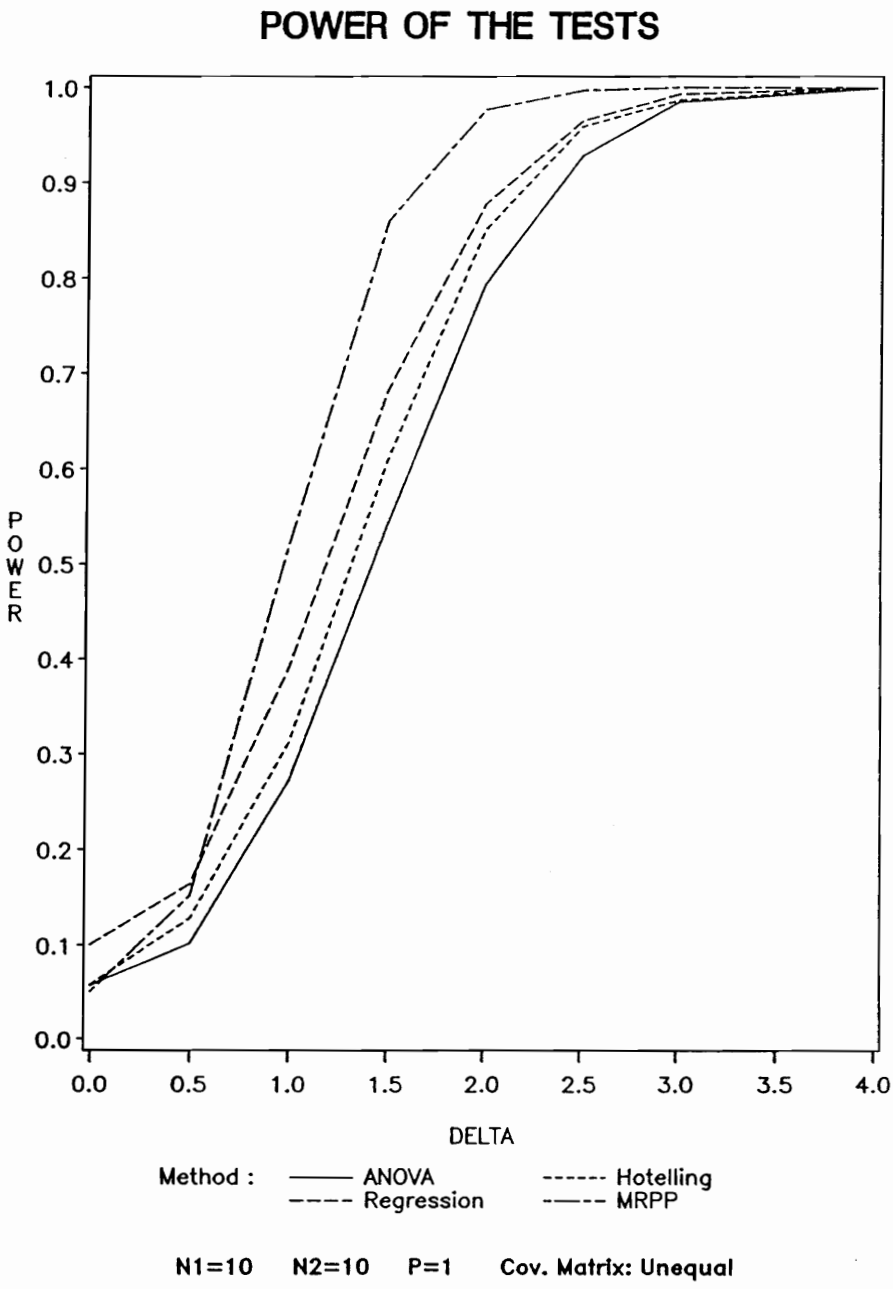

Figure 2.21. Power of the two-sample ANOVA, Regression, Hotelling T^2 and MRPP methods with no correlation present, covariance matrices unequal, sample sizes of $N_1=10$ $N_2=10$ and $P=1$.

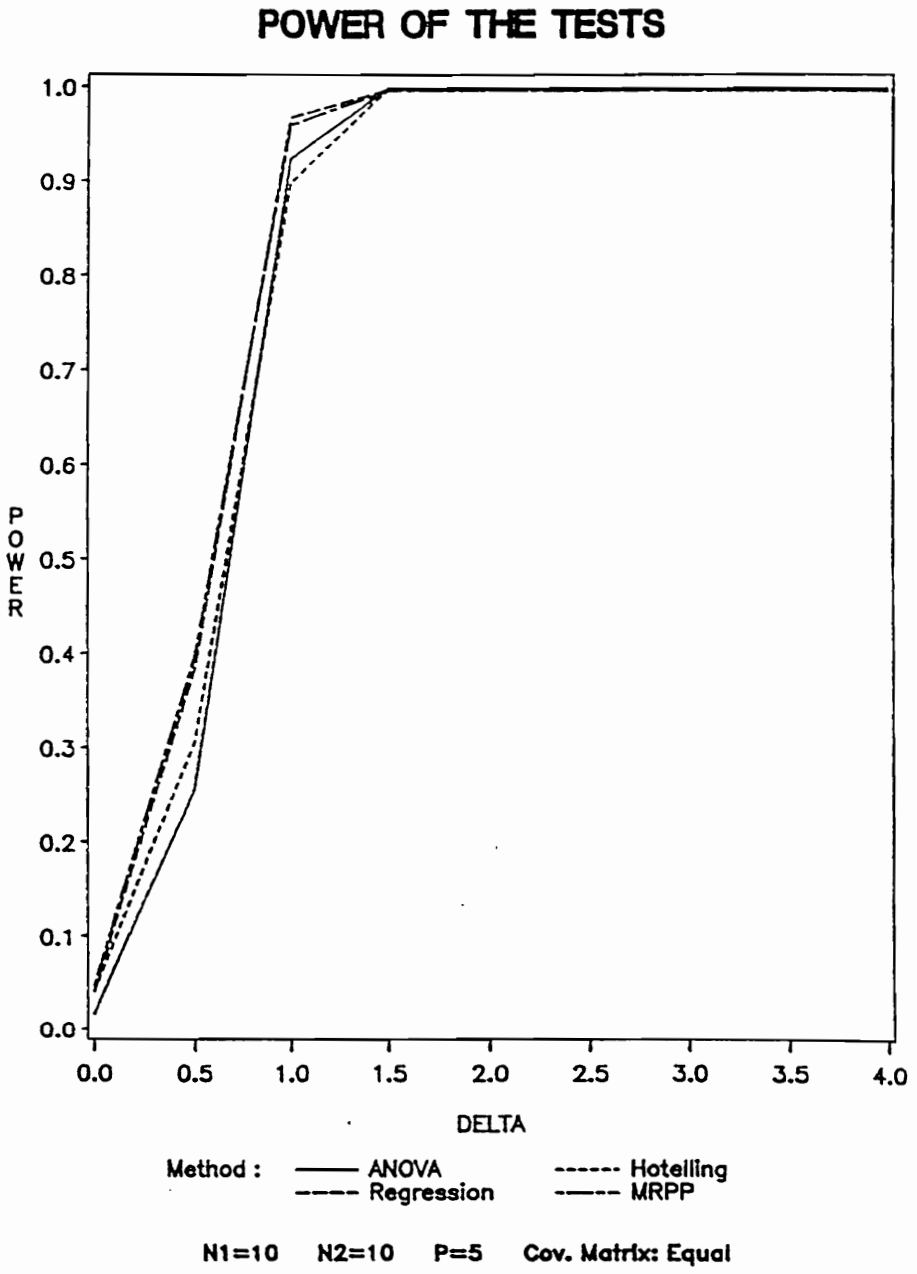

Figure 2.22. Power of the two-sample ANOVA, Regression, Hotelling T^2 and MRPP methods with no correlation present, covariance matrices equal, sample sizes of $N_1=10$ $N_2=10$ and $P=5$.

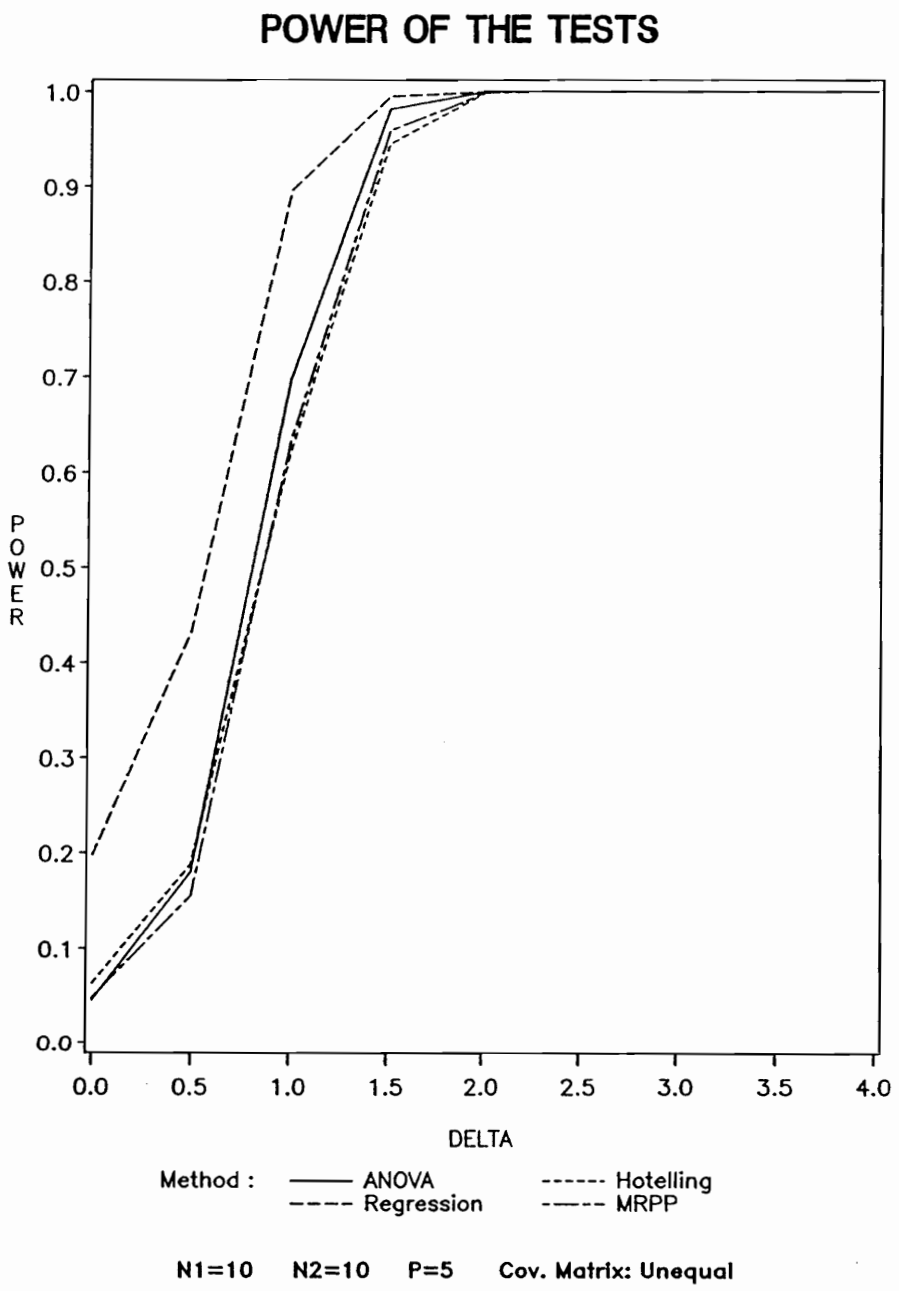

Figure 2.23. Power of the two-sample ANOVA, Regression, Hotelling T^2 and MRPP methods with no correlation present, covariance matrices unequal, sample sizes of $N_1=10$ $N_2=10$ and $P=5$.

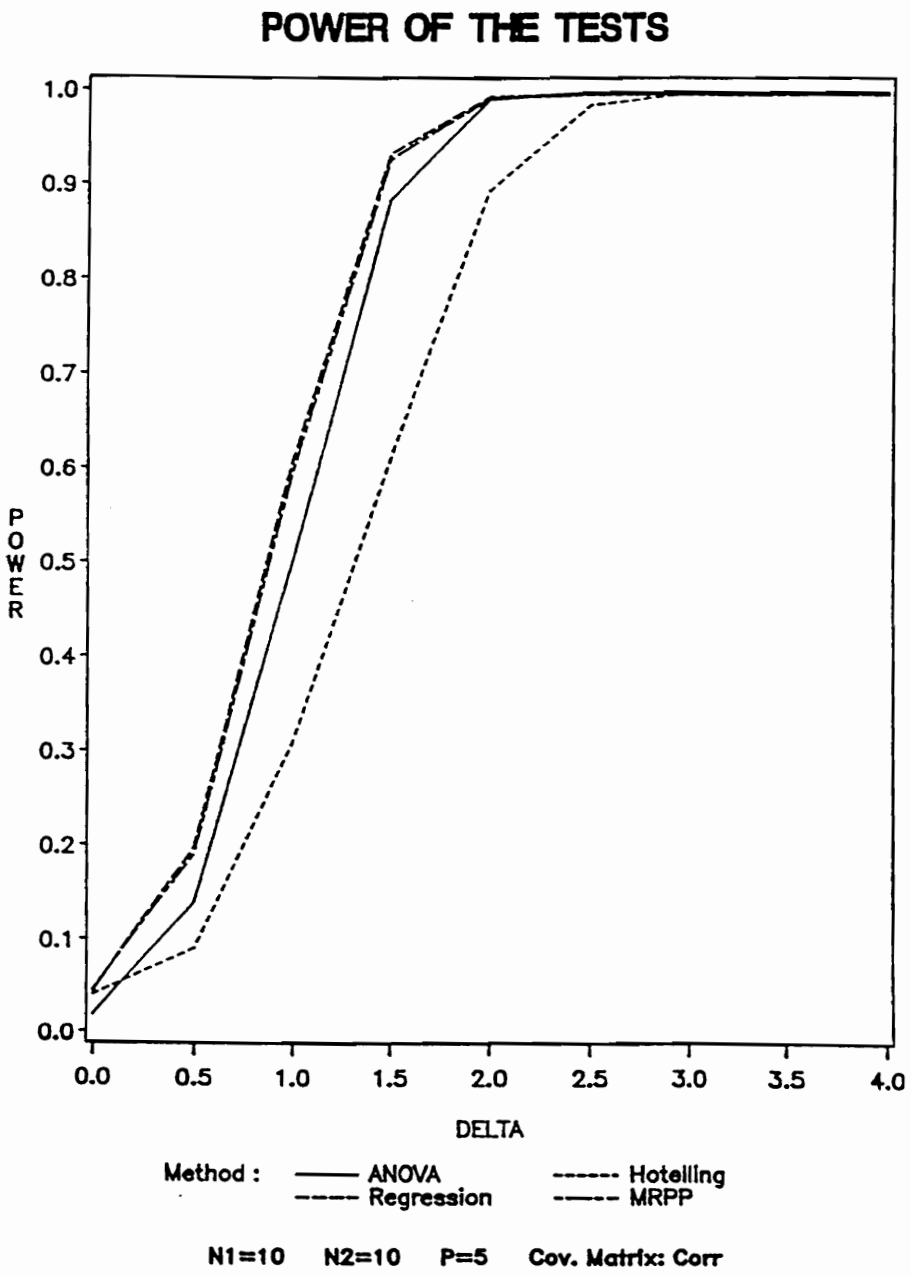

Figure 2.24. Power of the two-sample ANOVA, Regression, Hotelling T^2 and MRPP methods with correlation present (ρ=0.8), covariance matrices equal, sample sizes of N_1=10 N_2=10 and P=5.

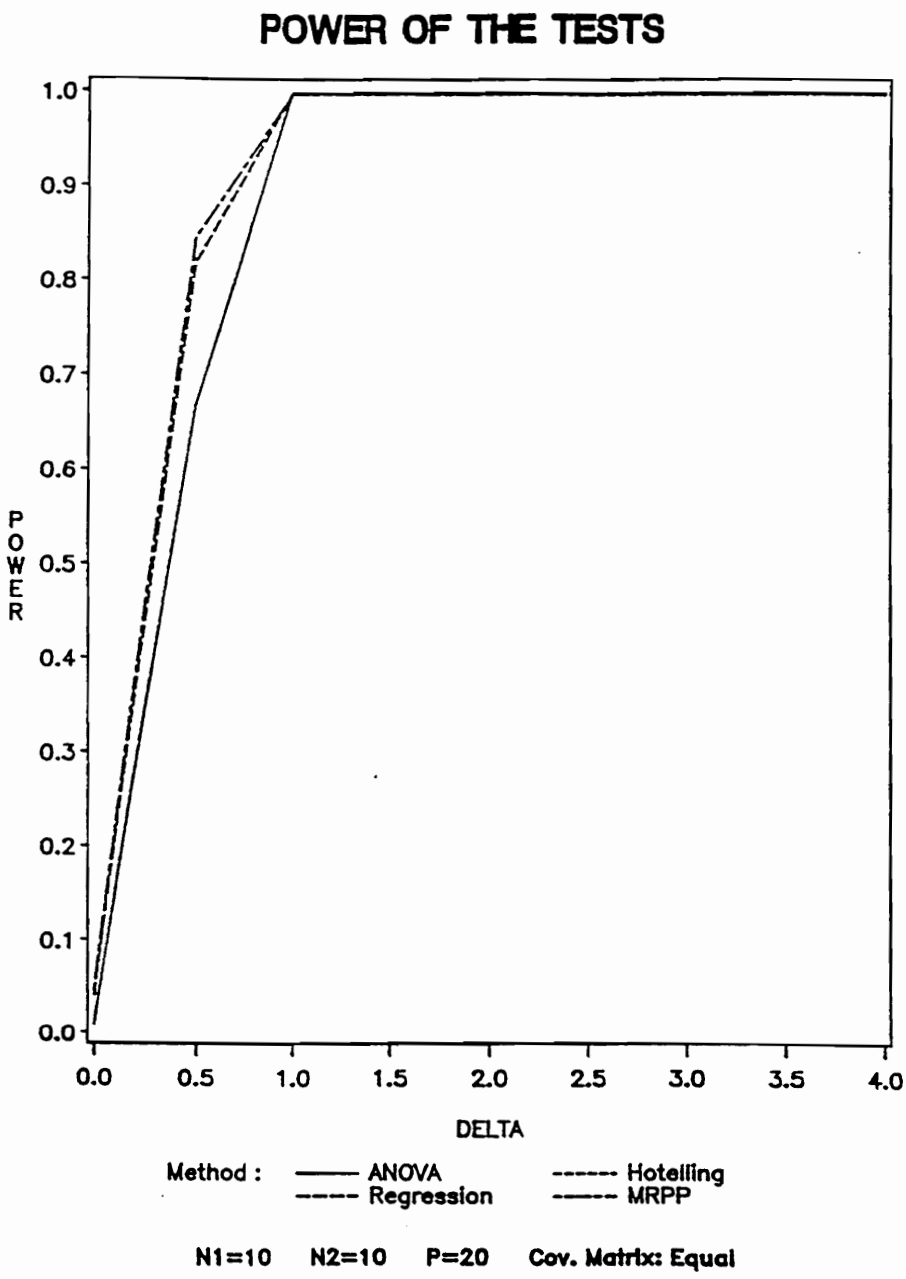

Figure 2.25. Power of the two-sample ANOVA, Regression and MRPP methods with no correlation present, covariance matrices equal, sample sizes of $N_1=10$ $N_2=10$ and $P=20$.

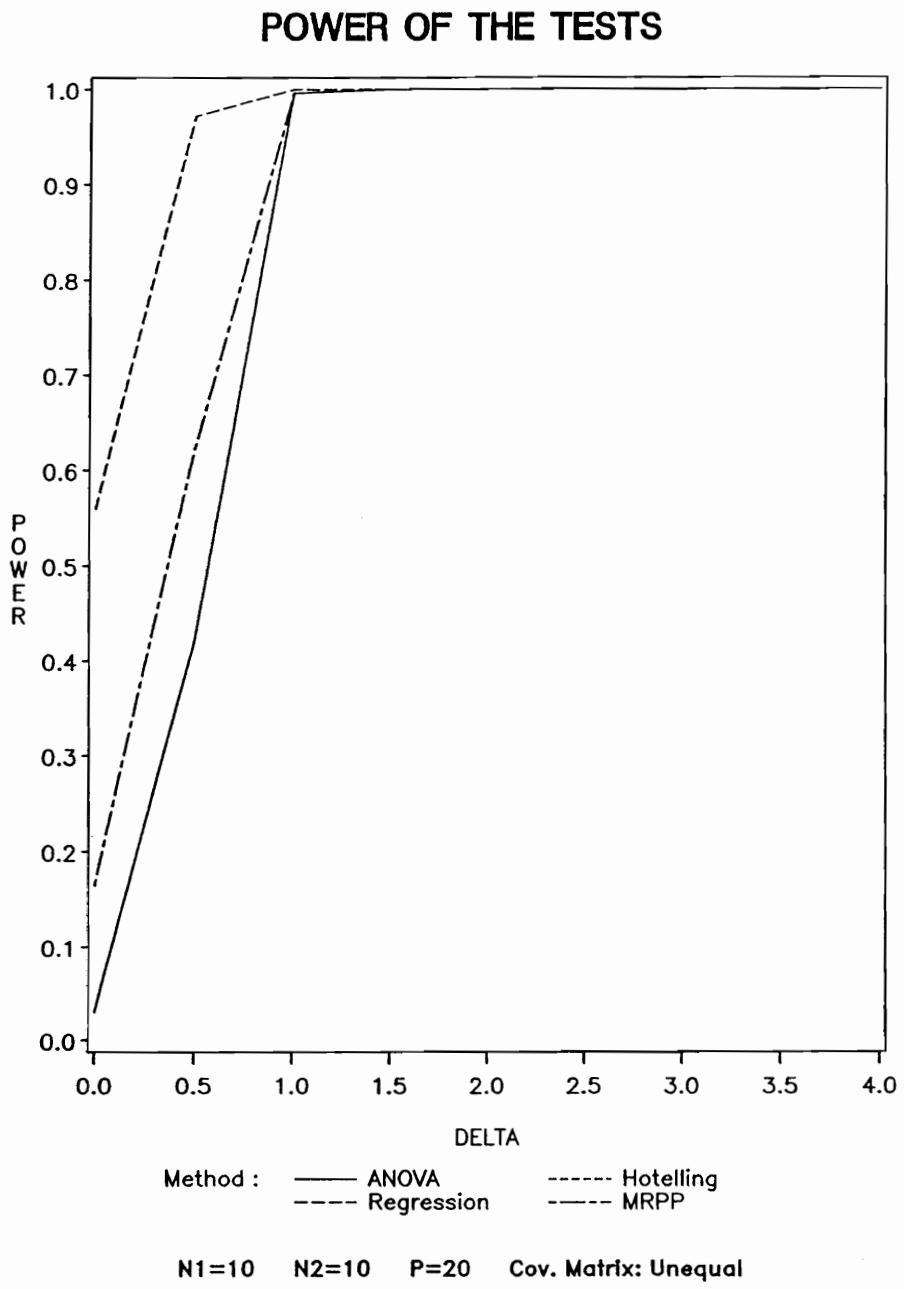

Figure 2.26. Power of the two-sample ANOVA, Regression and MRPP methods with no correlation present, covariance matrices unequal, sample sizes of N_1=10 N_2=10 and P=20.

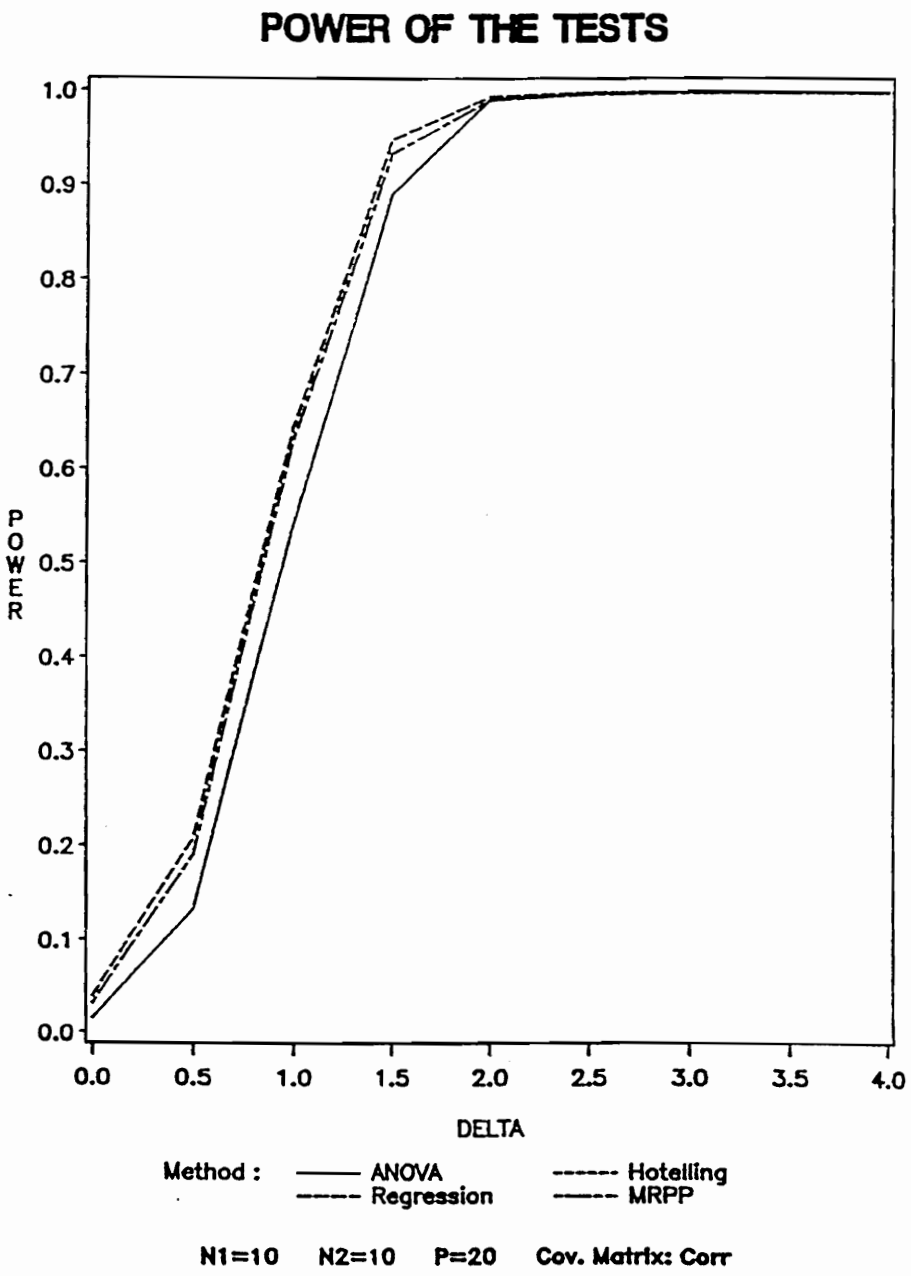

Figure 2.27. Power of the two-sample ANOVA, Regression and MRPP methods with correlation present (ρ=0.8), covariance matrices equal, sample sizes of N_1=10 N_2=10 and P=20.

§ 2.3.3.*C* *Effects of Data Rotation*

Figures 2.28 and 2.29 illustrate the impact of rotating one set of points $0-2\pi$ radians in increments of $\frac{\pi}{8}$ radians cyclically in the plane about a second set of points. Minor fluctuations from flat power profiles were noted when using data generated from the equal ($\Sigma_x=\Sigma_y$) and unequal ($\Sigma_x\neq\Sigma_y$) covariance structures. While similar power profiles were observed under the two covariance structures, power averaged 30% less when $\Sigma_x\neq\Sigma_y$ than when $\Sigma_x=\Sigma_y$.

A predictable cyclic power profile emerged when rotating sets of correlated data. Bimodal maxima and minima of the power profiles for the MRPP and regression procedures occurred at precisely the same locations. In fact, the two profiles were parallel with the regression profile translated above the MRPP profile by roughly 7%. Maxima occurred at $\frac{3\pi}{4}$ and $\frac{7\pi}{4}$ and minima at $\frac{\pi}{4}$ and $\frac{5\pi}{4}$ indicating that the power profiles, as a function of rotation, were sinusoidal in nature with period π. Due to the sinusoidal pattern, power averaged about the same using correlated data as when using data generated by the equal, uncorrelated covariance structure.

As one may surmise, data rotation appears to impose little effect on power provided the data are uncorrelated. Significant effect on the power profile derived from correlated data manifests itself in the form of a sinusoidal pattern of period π.

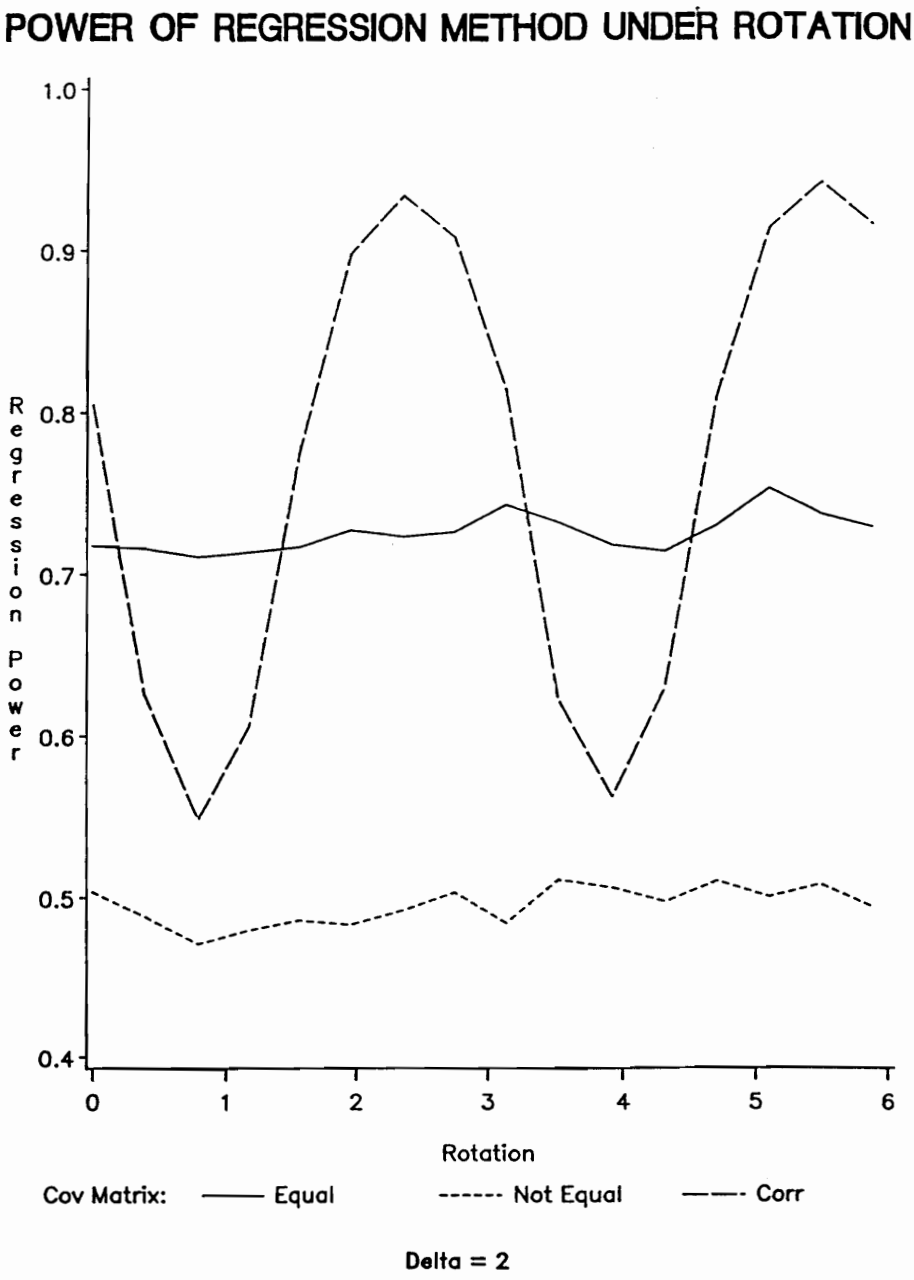

Figure 2.28. Impact on power of the regression method of rotating one set of points $0-2\pi$ radians in increments of $\frac{\pi}{8}$ radians cyclically in the plane around the second set of points.

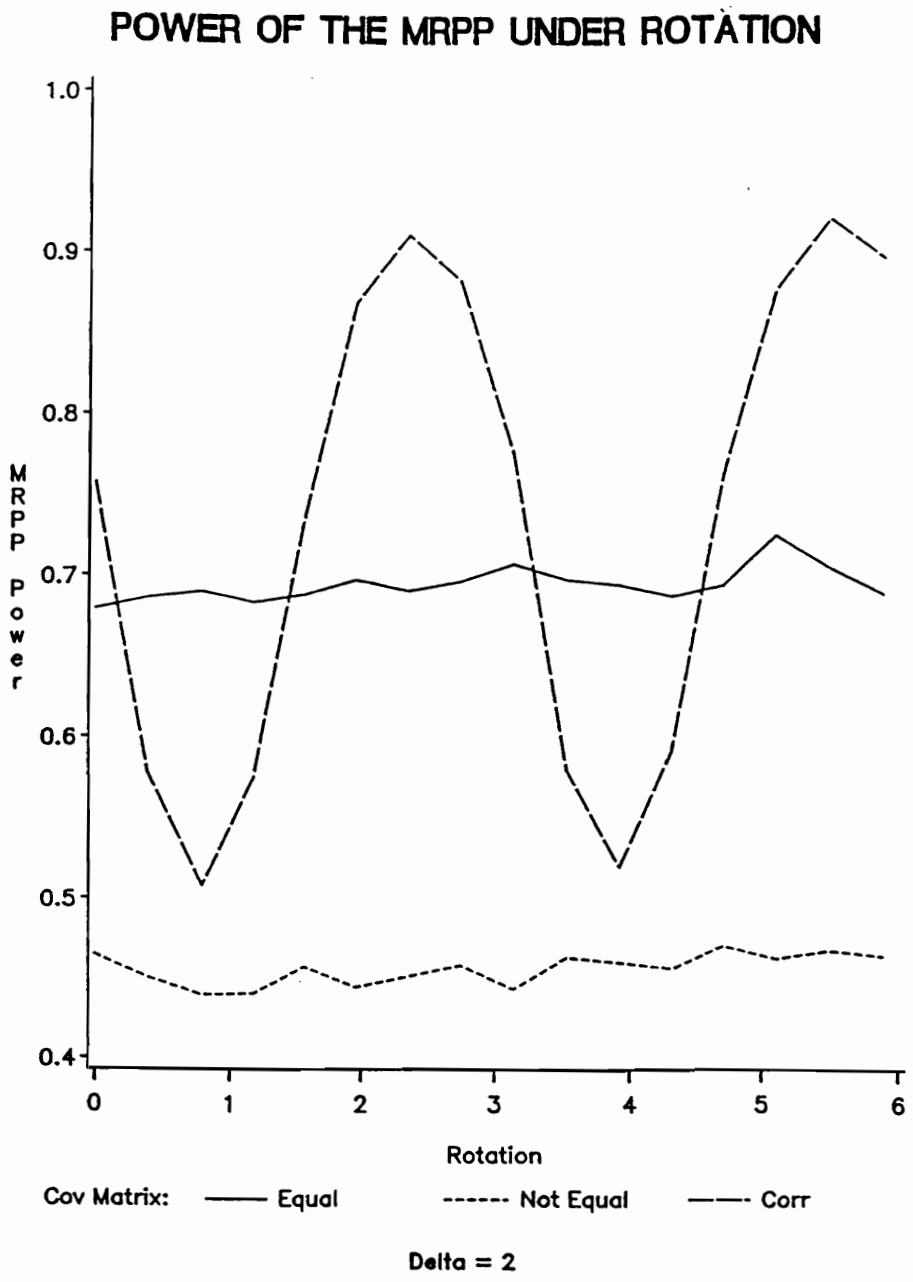

Figure 2.29. Impact on power of the MRPP method of rotating one set of points $0-2\pi$ radians in increments of $\frac{\pi}{8}$ radians cyclically in the plane around the second set of points.

§ 2.3.3.*D* *Outliers*

In keeping with an overall effort to assess the comparative robustness of the MRPP and regression methods to some common maladies afflicting observed data, this small simulation study was conducted to explore the effect of outliers on test performance.

Figures 2.30 and 2.31 illuminate the effect on power of an outlying observation on the regression method when the centroids for X and Y are separated by one and two standard deviations, respectively (δ=1, 2). As can be seen from Figures 2.32 and 2.33, results for the MRPP are virtually identical and the same discussion for the regression method applies to the MRPP. The outlier was formed by translating a single point in Y, denoted Y_i, by a fixed amount ζ=4 in all four orthogonal directions using the original point in Y as the origin. Presented in the figures are the results when ζ=4, similar results (not presented) were produced using a less extreme outlier with ζ=2.

Maximum power was achieved when the outlier was created by translating Y_i by $(\zeta,0)$ since this effected the greatest separation between X and Y (refer back to Figure 2.2). Power decreased to just below the level observed when no outlier was present when $(0,\zeta)$ was added to Y_i. Minimum power occurred after adjusting Y_i by $(-\zeta,0)$, as this outlier resulted in the minimum observed separation between X and Y.

Although it is not easily seen from Figure 2.30 ($\delta=1$) what role the covariance structure plays in the power profile when influenced by the outliers chosen, it becomes evident from Figure 2.31 ($\delta=2$) that the covariance pattern does influence test performance. The additional separation in the centroids of X and Y when $\delta=2$ resulted in better resolution of the role the correlation pattern plays in determining power characteristics. Power was greatest when using correlated data. A decrease of roughly 5% was observed when using equal variance, uncorrelated data and 15% when data were generated with no correlation and unequal variances.

Although the differences observed in performance were small under the three covariance structures it does indicate that the regression method is effective in modelling correlation while remaining particularly sensitive to heterogeneity.

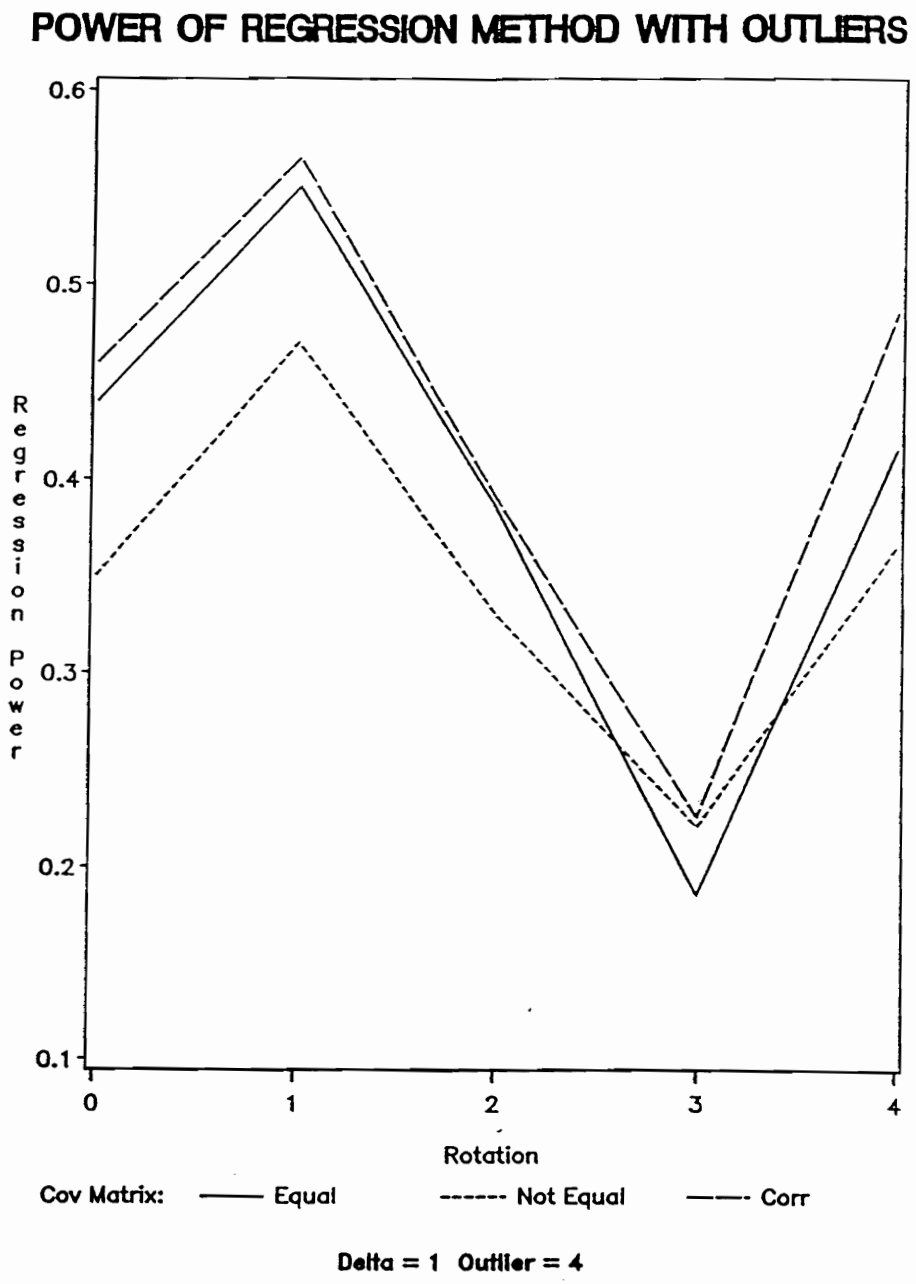

Figure 2.30. Effect on power of an outlying observation on the regression method when the centroids for X and Y are separated by one standard deviation (δ=1).

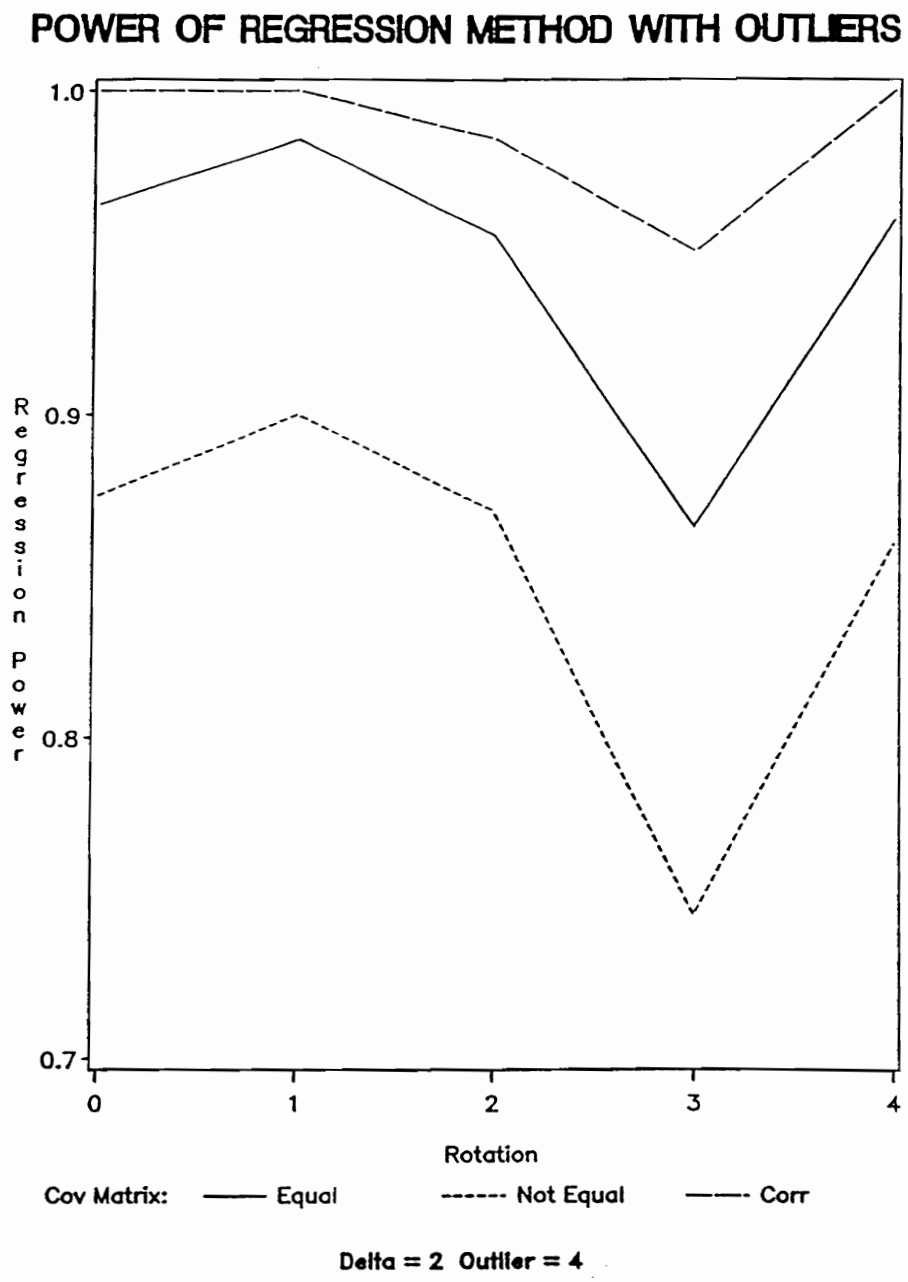

Figure 2.31. Effect on power of an outlying observation on the regression method when the centroids for X and Y are separated by two standard deviations (δ=2).

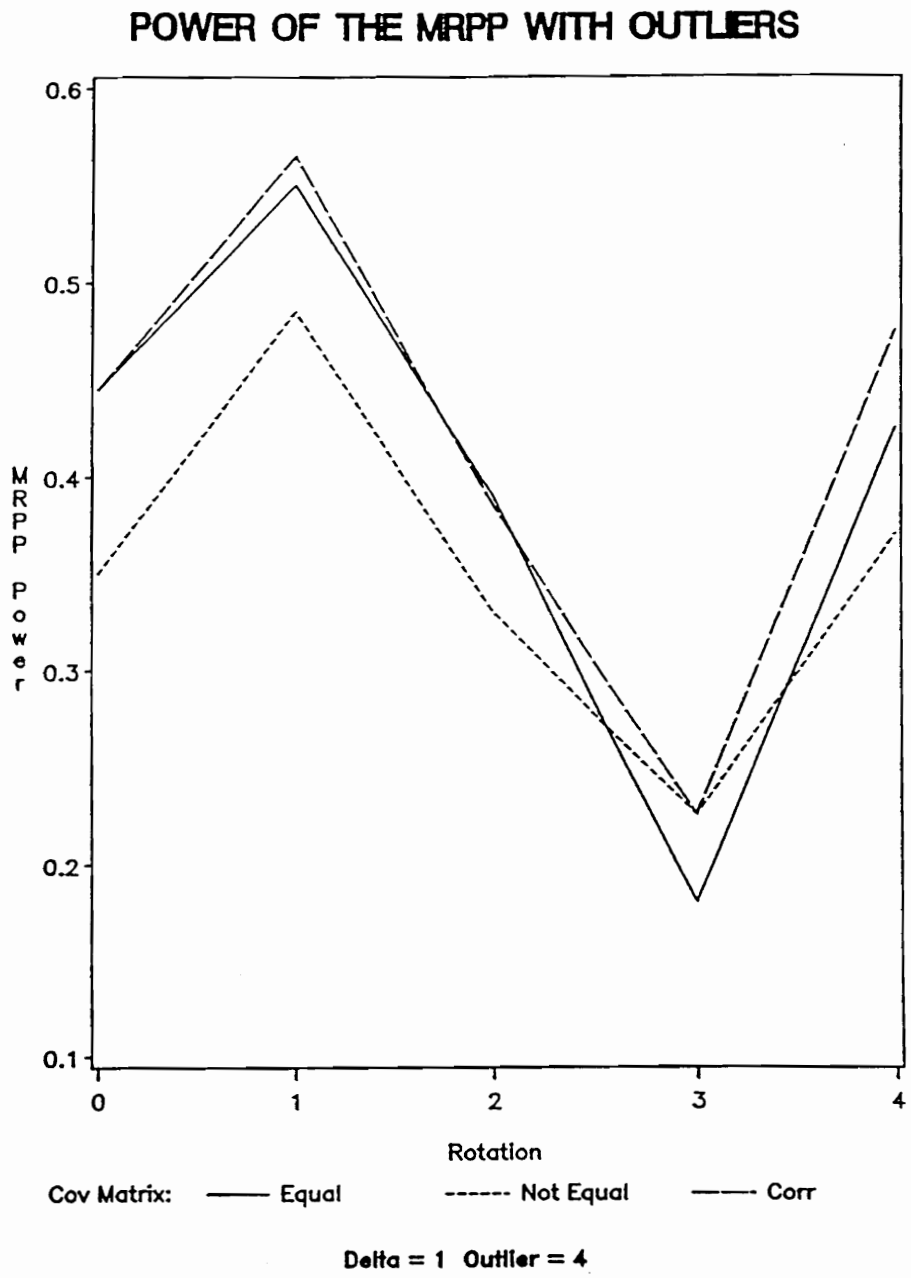

Figure 2.32. Effect on power of an outlying observation on the MRPP when the centroids for X and Y are separated by one standard deviation (δ=1).

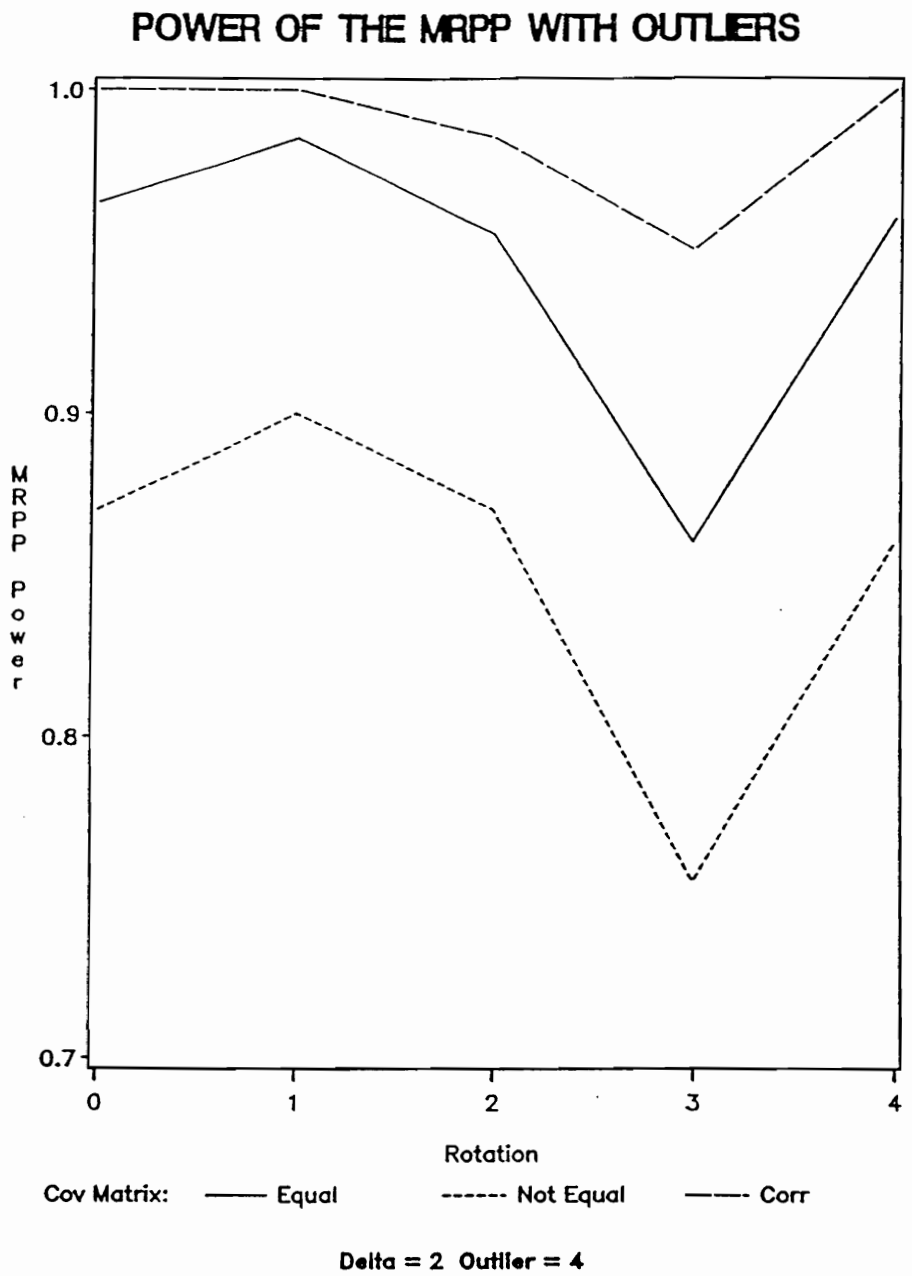

Figure 2.33. Effect on power of an outlying observation on the MRPP when the centroids for X and Y are separated by two standard deviations (δ=2).

§ 2.3.3.*E* *Effects of Non-Normality*

In this final small-scaled simulation study two independent samples of data, $X=(X_{j1},\ldots, X_{jP})$ and $Y=(Y_{j1},\ldots, Y_{jP})$ for $j=1,\ldots,10$ and $P=1, 2$, were generated following lognormal distributions given by $X_{ij} \sim LN(\exp(0.5),\ e(e-1))$ and $Y_{ij} \sim LN(\exp(\mu+\sigma^2/2),\ \exp(2\mu+2\sigma^2)-\exp(2\mu+\sigma^2))$, where the underlying distributions are given respectively for X_{ij} and Y_{ij} by $N(0,1)$ and $N(\mu,\sigma^2)$. It is likely to encounter data arising from multispecies microcosm experiments which do not follow normal law. Procedures used in the analysis of data resulting from these experiments should exhibit some degree of robustness to departures from normality.

Our efforts here focused both on the ability of the MRPP and regression procedures to hold true the α—levels and to insure adequate test performance against a series of alternative hypothesized values ($\delta=0.0, 0.5,\ldots, 3.0$). Only the equal and unequal covariance cases were involved in generating the data since for P=1 there can be no correlation. α—levels were held to the true 0.05 level under the equal covariance structure. When data were simulated with unequal covariances, α—levels doubled in the P=1 case and increased approximately 12-fold to 0.60 when P=5 (Figures 2.34-2.37).

Decreases in power of approximately 7% for the MRPP and 15% for the regression procedure were observed when generating data with unequal variances (P=1). Severely inflated α—levels (power at $\delta=0$) made direct comparison between

the power profiles of data generated by the two covariance structures difficult, but did indicate that both procedures are quite vunerable to non–normal data in this setting.

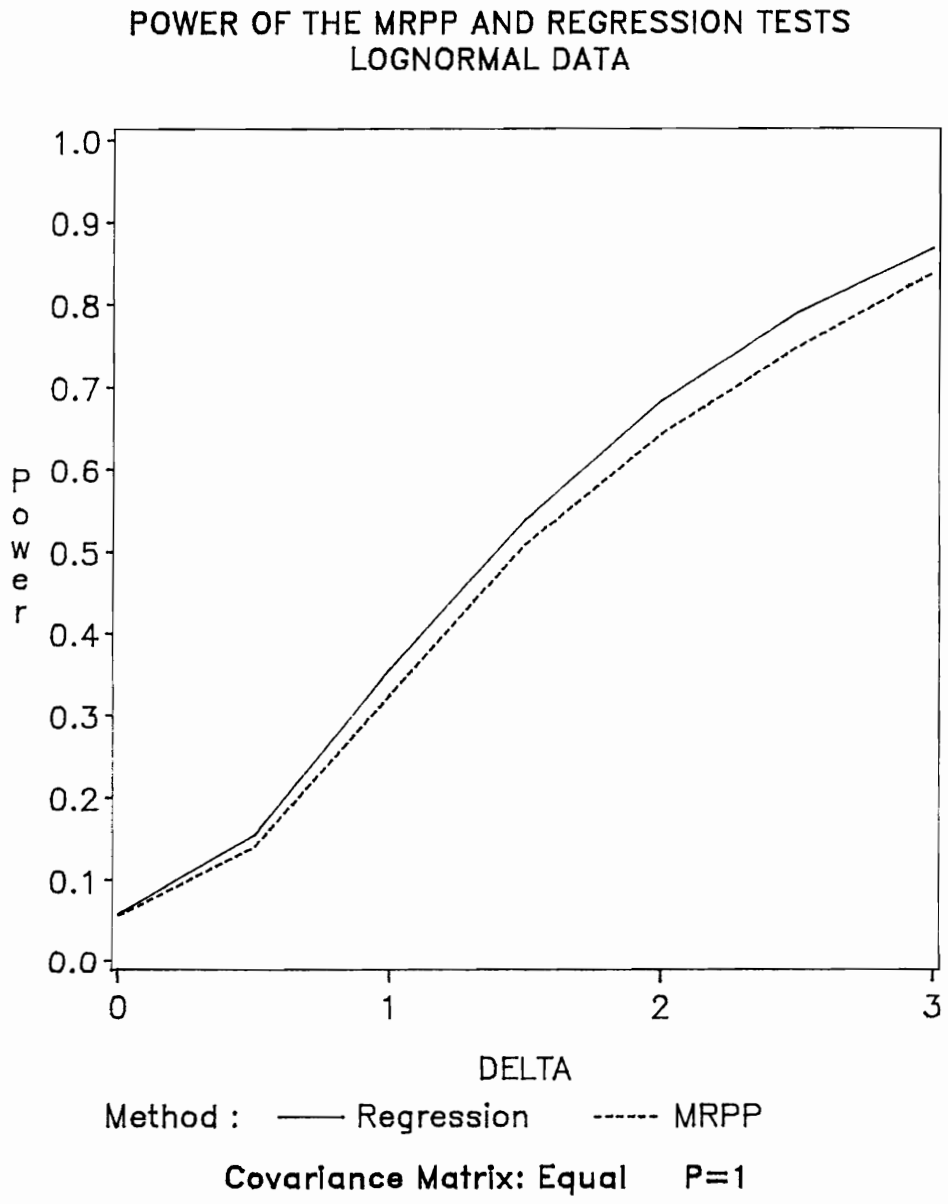

Figure 2.34. Effect on power of the MRPP and regression methods of lognormal data when the covariance matrices are equal and P=1.

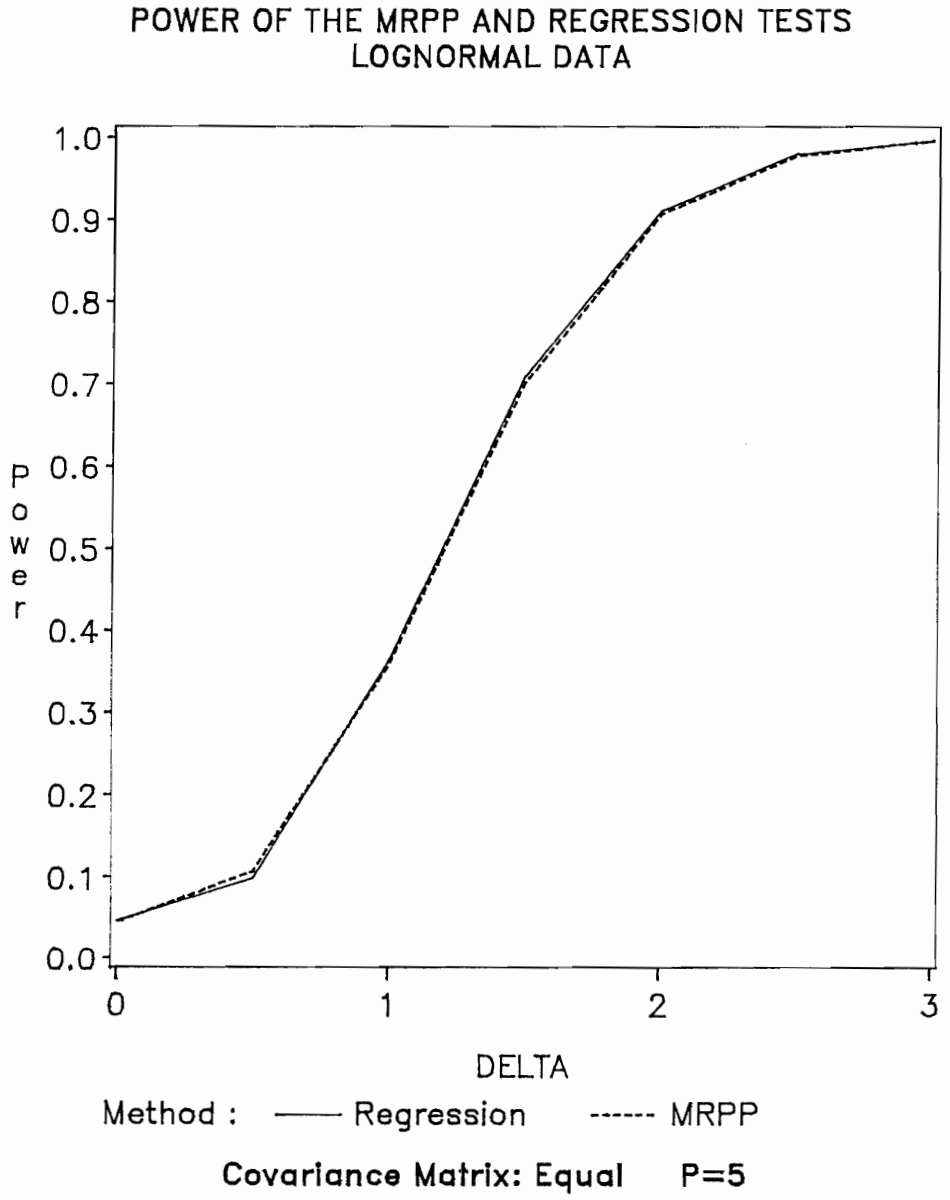

Figure 2.35. Effect on power of the MRPP and regression methods of lognormal data when the covariance matrices are equal and P=5.

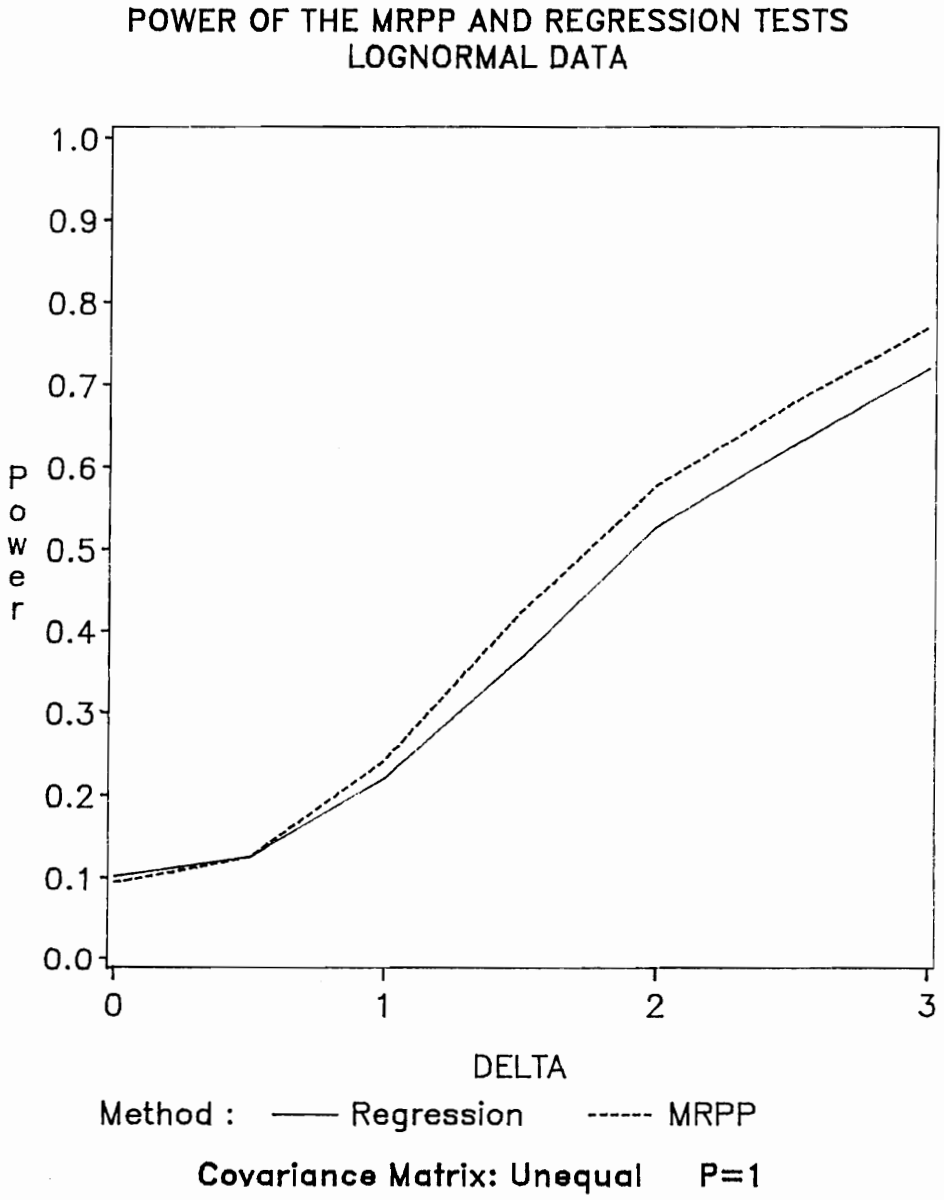

Figure 2.36. Effect on power of the MRPP and regression methods of lognormal data when the covariance matrices are unequal and P=1.

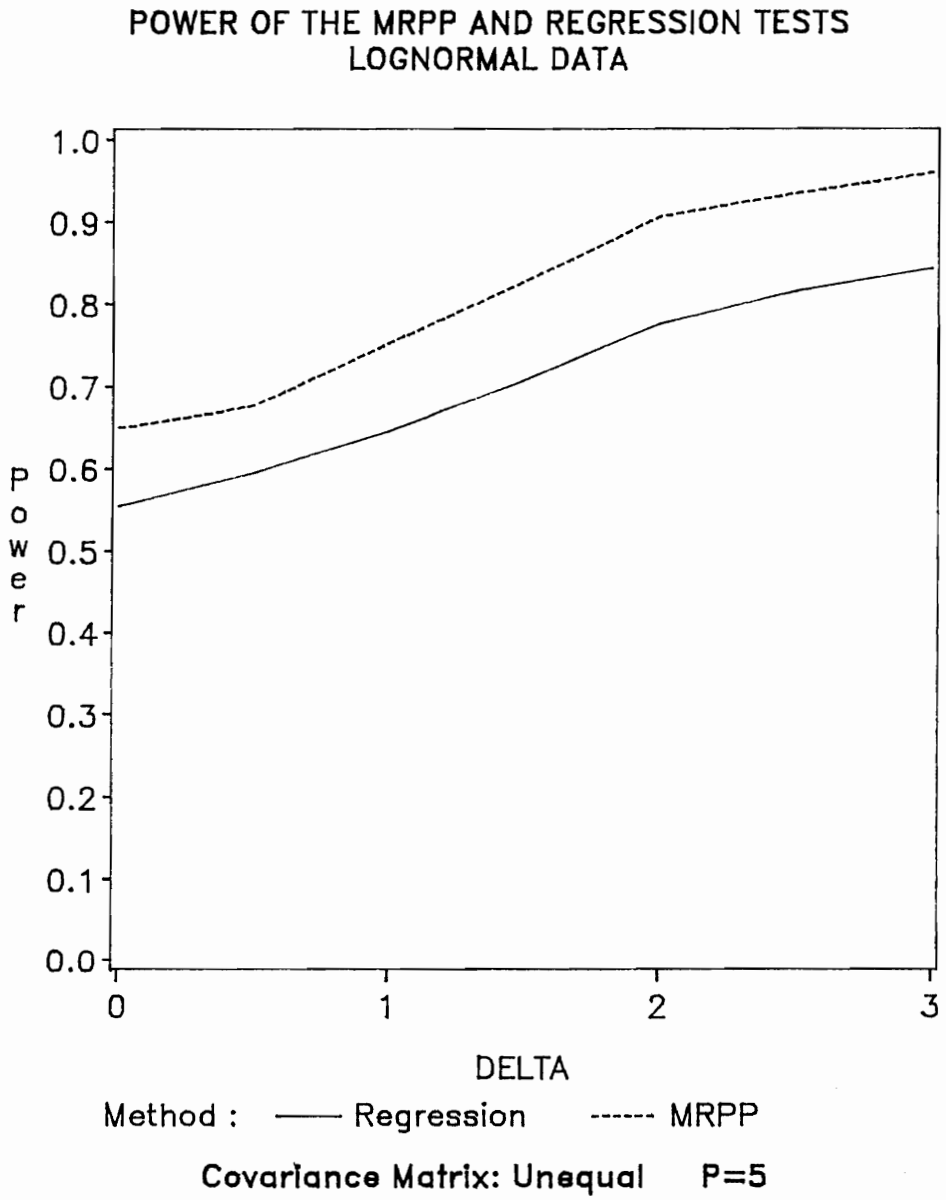

Figure 2.37. Effect on power of the MRPP and regression methods of lognormal data when the covariance matrices are unequal and P=5.

§ 2.3.3.*F* *Summary*

In general, power performance was best for all tests under the equal, uncorrelated covariance structure ($Cov(X) = Cov(Y) = I_P$). Noticeable diminishing of power occurred when using the unequal, uncorrelated covariance structure $Cov(X) = I_P$, $Cov(Y) = 3 \cdot I_P$. Further reduction in power resulted when using the correlated covariance structure $Cov(X) = Cov(Y) = I_P + \rho \cdot (J_P - I_P)$. Increases in total sample size and the number of dependent variables resulted in significantly better power characteristics.

The MRPP and regression method out-performed the remaining procedures in most circumstances considered. The MRPP would be more useful in experiments with replication and/or when normality of the responses is not normally distributed. The regression method would be used mostly in unreplicated experiments and/or the magnitude of treatment differences are of importance. The regression procedure can also be formulated in the framework of a permutation test, providing additional flexibility.

Data orientation through rotation had only slight effect on test performance when data were uncorrelated, but induced pronounced predictable periodic behavoir in the power profile when data were correlated (ρ=0.8). When considering the effects the outlier had on the two procedures it appeared that power was altered solely as a function of the resulting separation between X and Y. Both procedures performed

well against reasonable alternatives when the underlying distribution was non-normal (lognormal) as long as the variances for the two groups were equal. Possibly, data generated with different covariance structures may exhibit more α-level inflation.

CHAPTER III

SELECTION OF VARIABLES

§ 3.1 *INTRODUCTION*

In previous sections we have described a number of methods useful in assessing changes in a community due to an external agent such as a toxicant. The MRPP and regression methods were found to be superior to others in detecting differences between treatments. Additional studies of these two methods shed light on how performance was affected by such maladies as outliers, non-normality and redundant variables.

Interest now shifts to identifying species which contribute the greatest amount of information for detecting treatment differences. As is often the case in microcosm experiments, sample sizes (number of replicate microcosms) are small and dimension (number of species or dependent variables) large. Information regarding treatment differences is rarely evenly spread across all species, but is often concentrated in a few key species. It would be very useful to develop methods for selecting and retaining

those species which significantly contribute to the assessment of group differences and discarding those deemed redundant.

Potential applications of such methodology are wide-ranging, limited not only to the analysis of toxicological microcosm experiments but to many fields where multivariate data naturally arise. In particular for microcosm experiments, a savings of time and expense can be realized if the number of species is significantly reduced while at the same time retaining the integrity of the inferences made.

§ 3.2 *MEASURE OF IMPORTANCE*

Smith (1986) devised a method for evaluating the contribution (or importance) of individual species in detecting differences indicated by tests of location, such as the MRPP and regression methods. Formulated in the same spirit as many well known regression diagnostics, the method involves removing one species at a time and computing the quantity

$$I_r = 100 \times \frac{(\overline{B}_{-r} - \overline{B})}{\overline{B}}, \qquad 3.1$$

where $\overline{B}_{-r}$ is the mean between dissimilarity (similarity) with the r^{th} species removed and $\overline{B}$ is the mean between dissimilarity with all species included. I_r is the percentage relative importance of species r on the mean between dissimilarity.

It is important to note that the type of metric used may have a profound effect on the the importance measure, I_r. Smith (1986) used proportional abundance, rather than total abundance, and the metric given previously as equation 2.2. This metric was first proposed by Stander (1970) and is equivalent to the cosine of the angle between the vectors x_i and x_j and can be alternately expressed as

$$\mathrm{Cos}\theta_{x_i x_j} = \frac{\langle x_i, x_j \rangle}{\|x_i\| \|x_j\|}. \qquad [3.2]$$

Unlike the Euclidean distance-based metrics, the cosine metric is a similarity measure. It measures how close replicates are to each other in P-dimensional space. When using the cosine metric, importance values (I_r) can be either positive or negative. When using proportions, and to a lesser extent for total abundance, large positive values indicate species important in determining treatment differences. Values near zero reflect unimportant species while large negative values of I_r usually indicate species that contribute to the between dissimilarities but are relatively unimportant in determining treatment differences (Smith 1986).

Hotelling's T^2, F-statistics from ANOVA and regression, and other methods derived from least squares are based on *squared* (possibly weighted) Euclidean distance, whereas the MRPP is based on Euclidean distance (L_2—norm). Use of (squared) Euclidean distance-based metrics with total abundances result in importance values that are of the same sign (negative when using $\overline{B}_{-r}-\overline{B}$ or positive for $\overline{B}-\overline{B}_{-r}$). Some sensitivity of the measure is lost since only the magnitude of I_r can be considered in this case.

To show that all I_r have the same sign, let the importance of the r^{th} species be defined as

$$I_r=\frac{\overline{B}-\overline{B}_{-r}}{\overline{B}}. \qquad [3.3]$$

Let squared Euclidean distance be the metric, the average between dissimilarity without and with the r^{th} species removed can be expressed as

$$\bar{B}=\sum_{k=1}^{g}\frac{1}{M_K}\left(\sum_{i=S_1}^{S_2}\sum_{j=S_3}^{N}B_{ij}\right)$$

$$=\sum_{k=1}^{g}\frac{1}{M_K}\left\{\sum_{i=S_1}^{S_2}\sum_{j=S_3}^{N}\left(\sum_{m=1}^{P}(X_{im}-X_{jm})^2\right)\right\}, \qquad [3.4]$$

$$\bar{B}_{-r}=\sum_{k=1}^{g}\frac{1}{M_K}\left[\sum_{i=S_1}^{S_2}\sum_{j=S_3}^{N}\left[\sum_{\substack{m=1\\m\neq r}}^{P}(X_{im}-X_{jm})^2\right]\right], \qquad [3.5]$$

$$\bar{B}-\bar{B}_{-r}=\sum_{k=1}^{g}\frac{1}{M_K}\left\{\sum_{i=S_1}^{S_2}\sum_{j=S_3}^{N}(X_{ir}-X_{jr})^2\right\}\geq 0 \qquad [3.6]$$

where

$$M_K=\sum_{i=k+1}^{g}n_i n_k,\quad S_1=\sum_{l=1}^{k}n_{l-1}+1,\quad S_2=\sum_{l=1}^{k}n_l,\quad S_3=\sum_{l=1}^{k}n_l+1,$$

g is the number of treatment groups and n_0 is defined as zero. The importance measure $I_r\geq 0$ since the numerator and denominator are both ≥ 0 and equals zero only if $x_{ir}=x_{jr}$, $\forall$ i and j. I_r is the ratio of how much each replicate differs from the others on the r^{th} species to how much the replicates differ over all species.

One advantage of using squared Euclidean distance is that the proportion of the total between dissimilarity attributable to each variable (species) can be easily computed since the sum of importance values for all species is unity, i.e.,

$$\sum_{r=1}^{P} I_r = 1.$$

This is evident from equations 3.4 and 3.6.

A new, crude diagnostic is now developed for importance values to aid in the selection of important species. Assume $(X_{ir}-X_{jr}) \sim N(0,\sigma^2)$ then $(X_{ir}-X_{jr})^2 \sim \sigma^2\chi_1^2$ which is the squared Euclidean distance between the i^{th} and j^{th} replicates for the r^{th} species. If we sum over all species we have

$$\sum_{j=1}^{p} (X_{ir}-X_{jr})^2 \sim \sigma^2\chi_p^2. \qquad [3.7]$$

From equation 3.4,

$$\frac{1}{M_K}\left\{ \sum_{i=S_1}^{S_2} \sum_{j=S_3}^{N} \left(\sum_{m=1}^{P} (X_{im}-X_{jm})^2 \right) \right\}$$

represents the average squared Euclidean distance for the k^{th} group which is roughly distributed as $\sigma^2\chi_p^2$. Summing over the g groups this approximation becomes $\sigma^2\chi_{gp}^2$. Likewise it follows that the distribution of eq. 3.6 roughly follows a $\sigma^2\chi_g^2$ distribution. Combining these results we have

$$\frac{\overline{B}-\overline{B}_{-r}}{\overline{B}} \doteq \frac{\sigma^2\chi_1^2}{P\sigma^2\chi_p^2/P} \sim \frac{1}{P} F_{g,g\cdot p}. \qquad [3.8]$$

We can use the diagnostic, $\frac{1}{p} F_{g,g\cdot p}$, as a crude approximation to the cut–off value for importance. The reader should note that the numerator and denominator are χ^2–like and will not, in general, be independent. Hence caution should be exercised in applying this measure too stringently. This diagnostic will be applied later to data from an experiment investigating the role of bacteria in gingivitis.

§ 3.3 *STEPWISE DEPENDENT VARIABLE SELECTION ALGORITHM*

The importance measure just discussed has many attractive features for use in dependent variable selection. Its ease of computing and interpretation are particularly noteworthy. An apparent weakness of the measure is that it considers only the full model. For instance, it does not address the question of whether a variable remains important in explaining the differences observed between treatments when other variables are already present. New methods are presented in the remainder of this chapter which address this apparent shortcoming.

A stepwise ***dependent*** variable selection algorithm is proposed that will work with many types of selection criteria. As in the classical regression stepwise variable selection procedure, a criterion for entry or exit of variables is required. A criterion is proposed that will enable objective evaluation of a variable's contribution in assessing treatment differences. The criterion is permutation-based and, at each step, looks at the conditional p-values obtained for each variable not yet retained. It then retains that variable with the lowest or most significant p-value given that other variables may already be kept.

Forward selection and backward elimination techniques are briefly considered as well. An additional criterion which is easily computed, distance-based, and non-permutational is proposed. The criterion enables variable selection based on maximizing (minimizing) $\overline{B}$ for the forward (backward) technique. Variable selection

using importance values is effected through the use of the crude diagnostic.

§ 3.3.A The Algorithm – ANOVA Setting

As a first step, the unconditional p-values are obtained separately for each dependent variable (species). Select as the first variable for retention that which produced the smallest p–value below some entry threshold level (e.g., ≤ 0.10). Data are permuted for each of the remaining $P-1$ variables, one at a time, while holding fixed the data for the variable retained. This results in $P-1$ conditional p–values. Select as the second variable that which has the smallest p–value of the remaining $P-1$ variates, conditioned on fixing (not permuting) the data for the variable retained, but below the entry threshold level. At this point, the order of the variables should be reversed. A test is conducted to determine if the first variable retained should still be kept when the second variable is considered fixed. We permute data for the first variate and compare the resulting p–value against a more liberal exit level (e.g., ≤ 0.20).

This cycle of fixing the data for variables retained, obtaining conditional p–values for the remaining variables, retaining the variable associated with the lowest p–value (subject to entry threshold) and re–examining those variables previously retained is repeated until no new variables can be retained. Many factors can influence the order in which variables are retained. Selection criteria and the metric determine in large

measure the ordering.

A couple of notes are worth mentioning here. If X_i and X_j are independent, $\forall$ i and j, then $Corr(X_i, X_j)=0$ and the p-values previously referred to as conditional are in reality unconditional bonafide p-values. Testing variables for exit would not be necessary in this case. Secondly, the algorithm as described above would need modification to work if the data were proportions. Minimally, we would need to start with two species.

The next section provides details on the proposed dependent variable selection criterion appropriate for data arising from microcosm experiments. No doubt many more criteria could be effectively employed. For example, instead of using the p-value from the MRPP, we could choose variables (species) which produce the largest univariate t-statistics. Alternatives to the stepwise algorithm are briefly discussed in subsequent sections. The chapter concludes with an application of the stepwise and importance methodology to a real data set taken from an experiment investigating the role of bacteria in gingivitis.

§ 3.3.B The Selection Criterion

Recall from §2.1.4 that the MRPP test statistic,

$$\delta = \sum_{i=1}^{g} C_i f_i$$

is the weighted average of the between dissimilarities (distance function). To determine the exact p–value for the MRPP requires that complete enumeration of all possible permutations of the data be performed. Then under the null hypothesis of a completely random allocation of the N population elements in Ω to the g subgroups S_i of size n_i corresponding to the *a priori* specified subgroups, S_i, we have M equally probable values for the test statistic δ, each with probability $\frac{1}{M}$, where M takes the value

$$M = \frac{N!}{\prod_{i=1}^{g+1} n_i!} .$$

The exact MRPP p–value (observed significance level) is the proportion of the δ's less than or equal to the observed δ value. In an α–level test, H_0 is rejected in favor of H_1 when the p–value is less than or equal to α. The first selection criterion is based on p–values calculated in this manner.

After the initial step in the stepwise dependent variable selection strategy, data on

one variable will be permuted while data for "retained" variables are held fixed. This gives us a much better idea of the importance of variables as some accounting is made for interrelationships between them.

§ 3.4 *FORWARD DEPENDENT VARIABLE SELECTION ALGORITHM*

An alternative to the stepwise algorithm and p–value criterion is presented that takes into account the contribution of individual species to the separation of treatment group centroids through the between dissimilarities. The algorithm is based on the widely known forward selection algorithm of ordinary regression. The selection criterion is to choose variables which maximize (when using a distance metric) the between dissimilarities. This is an intuitive criterion to use as we would ideally want replicates from the same treatment group to be similar to one another and replicates from different groups to be dissimilar. The $\max(\overline{B})$ criterion incorporates this idea.

Let $\overline{B}_{j|i,\ldots,k}$ represent the average between dissimilarity for the j^{th} variable given that variables i,...,k are already entered. As a first step, retain the variable which maximizes $\overline{B}$, say variable j. In the second step, we calculate $\overline{B}_{j|\cdot}$ for each variable along with the first retained variable, again choosing the variable maximizing $\overline{B}$. At each subsequent step we retain variables in the same manner. Unlike the stepwise algorithm, variables are entered but never discarded once retained.

As in the stepwise algorithm, an entry threshold level should be established *a priori* when using the p–value criterion. The p–value criterion may not be as sensitive when used with the forward (or backward) procedure as with the stepwise algorithm.

The max($\overline{B}$) criterion has no formal test, hence no p–value associated with it. Variable selection with this criterion can be accomplished by with a graphical aid. Plot relative cumulative max($\overline{B}$) versus step number. Look for an "elbow" or bend in the plot to determine the best number of important species or use some cut–off level of, say, 80% of the cumulative maximum ($\overline{B}$).

§ 3.5 *BACKWARD DEPENDENT VARIABLE ELIMINATION ALGORITHM*

Backward elimination of dependent variables works in an identical manner as for regressor variables in ordinary regression. We start with all variables present, eliminate one variable at a time and note the change in p–value or $\overline{B}$, depending on the criterion used. The criteria must be modified from the stepwise and forward selection procedures. At any one step, the variable with the *largest* p–value or *smallest* change in $\overline{B}$, indicating the least important variable, would be discarded subject to some exit level (e.g. 0.15 or 0.20).

Again, exit threshold level should be established *a priori* when implementing the p–value criterion. The graphical aid for use with the max($\overline{B}$) criterion would be equally suitable here.

Backward elimination would be useful in situations where many dependent variables are present and most were important in explaining treatment differences.

§ 3.6 *EXAMPLE: BACTERIA AND GINGIVITIS*

As an example, we take a subset of data from an experiment investigating the role of twelve species of bacteria in gingivitis, a common medical condition characterized by inflammation of the gingival tissue surrounding the teeth. Presented in Table 3.1 are species abundances, the number of bacteria observed in randomly chosen sections of gingival tissue. Notice how species 12 and 6 dominate the abundances. Table 3.2 lists the corresponding importance values using squared Euclidean distance, Euclidean distance and the cosine metric. Under the Euclidean distance–based norms species 12 appears to be by far the most important, followed by species 6. Using the cosine metric, species 12 is largest in magnitude but is assigned a negative value, possibly indicating that species contributes considerably to between dissimilarity (similarity) values but is not very important in explaining treatment differences.

Examination of importance alone would most likely lead one to consider species 12 and 6 only which is substantiated by the use of the crude diagnostic which yields a cut–off value of $3.4/12 \simeq 0.28$. Interestingly, the importance value for species 12 when using the cosine measure does down weight the importance in accounting for treatment differences. This is consistent with the interpretation of unconditional p–values listed in Table 3.3 which were generated by performing the MRPP (using Euclidean distance) with only one species at a time. P–values from Table 3.3 indicate that species 3, 6, 5 and to a lesser extent species 12 are the most important. These are the same p–values generated in the first step of the stepwise dependent variable

Table 3.1. Abundances for twelve species of bacteria with eight replicates and two treatment groups. Data taken from an experiment investigating the role of bacteria in gingivitis.

	Species											
Replicates	1	2	3	4	5	6	7	8	9	10	11	12
1	47	26	12	7	6	312	6	13	9	16	31	2655
2	14	44	2	1	14	292	32	31	9	16	32	1886
3	31	42	6	27	6	276	14	44	9	16	31	2071
4	42	22	4	3	20	290	9	30	6	8	18	2037
5	52	30	0	5	2	102	0	21	8	15	37	4046
6	12	22	0	4	4	50	16	20	11	16	30	2037
7	19	46	0	8	0	91	8	32	5	24	31	2658
8	86	46	0	11	6	97	13	4	4	6	15	3432

Table 3.2. Importance values from an experiment investigating the role of bacteria in gingivitis. Metrics used were Euclidean distance (E), squared Euclidean distance (E^2) and cosine (Cos).

	Species											
Metric	1	2	3	4	5	6	7	8	9	10	11	12
E	-.065	-.020	-.004	-.011	-.009	-5.72	-.009	-.024	-.001	-.004	-.004	-79.289
E^2	-.072	-.014	-.003	-.008	-.008		-.001	-.022	-.001	-.004	-.007	-96.914
Cos	.004	.003	.000	.001	.002	.596	.003	.005	.000	.001	.002	-13.820

Table 3.3. Unconditional p–values in tests (MRPP) of treatment differences considering each species individually. Data taken from an experiment investigating the role of bacteria in gingivitis.

	Species											
Criterion	1	2	3	4	5	6	7	8	9	10	11	12
P–Val.	.73	.70	.014*	.67	.071*	.014*	.99	.41	.24	.64	.90	.16*

* Species which appear important.

selection scheme. The Euclidean distance–based norms appear to lose the ability to discern between importance and magnitude since all importance values will be of the same sign.

P–values for the first three steps of the stepwise algorithm applied to the gingivitis data are presented in Table 3.4. Entering species 3 and 6 alone resulted in the same p–value (p=0.0143), consequently both were entered in the first step. Conditional testing of each species was performed by permuting data for that species and holding fixed data for the other species. In either case, species 3 and 6 remained important. Given that species 3 and 6 were retained in step one, species 5 was judged the most important (p=0.043) in step two. Again, each species contribution in explaining treatment differences was assessed in light of the new information received from species 5. P–values from all conditional tests indicated that species 3, 6 and 5 should be retained.

In the third and final step, the algorithm entered species 12 with a conditional p–value of p=0.100, which was the threshold used for retaining variables. After checking if any of the retained species could be dropped, given species 12 was now entered, the algorithm stopped. None of the species produced p–values below the threshold of p=0.10. The nearest was species 1 with p–value of p=0.200.

It is evident, at least for this example, that inferences regarding which species are important in detecting treatment differences based on importance values and the stepwise algorithm may lead to very different conclusions, regardless of the metric. In

Table 3.4. P-values resulting from applying the first three steps of the stepwise dependent variable selection algorithm to the gingivitis study data.

	Step											
Row	----1----			-----2-----			----------3----------					
E|F[1]:	6,3|--	3|6	6|3	5|3,6	6|3,5	3|5,6	12|3,6,5	5|3,6,12	6|3,5,12	3|12,5,6		
P-V.:	.014[2]	.00	.00	.04	.00	.00	.10	.00	.53	.07		

[1] F=Variable(s) fixed, E=Variable to be entered, the symbol E|F means E given F.
[2] Unconditional P-value for either species 3 or 6, all others are conditional.

an effort to shed more light on this problem the following comparison of hypothesis tests was conducted.

The MRPP was performed *removing* the most important species in the reverse order suggested by the species rankings from the importance values and stepwise algorithm, i.e., the most important species, the one retained first, would be the first to be deleted (Table 3.5). Next to be deleted in *addition* to the first variable would be the second most important species, and so on. The p–values at each step for both methods indicate that species 12 should be retained. Removing species 12 results in drastic increases in the p–values, making them highly insignificant.

Rather strong support for the stepwise algorithm is indicated in Table 3.5. If we remove the four species (3, 5, 6 and 12) selected as most important by the stepwise algorithm and run the MRPP, the resulting p–value is 1.00. This indicates all important species were removed and only unimportant species remain.

Table 3.5. P–values resulting from sequentially performing the MRPP on the gingivitis data while removing one additional species at a time (cumulative) as suggested in the rankings of species importance from the importance measure and stepwise algorithm using the L_2 norm.

MRPP

Importance Measure		Stepwise Algorithm	
Order	P–Value	Order	P–Value
6	0.138	6	0.138
12	0.830	3	0.138
8	0.840	5	0.138
1	0.680	12	1.000

CHAPTER IV

TESTING AND ESTIMATION IN REGRESSION

§ 4.1 *INTRODUCTION*

In the previous chapter we were primarily concerned with selecting a subset of variables or species that would retain most of the information contained in all the variables with respect to detecting treatment group differences. We now focus our attention on situations where regression is a more appropriate form of analysis.

Traditional regression analysis attempts to relate a group of independent or predictor variables to a response through a function or model which is linear in the parameters. Typical assumptions required for valid interpretation of model parameter estimates are independence of the errors resulting from fitting the model and regarding the levels of the regressor variables as fixed and under control of the researcher.

Throughout this dissertation we have been concerned with the analysis of data arising from mutispecies microcosm experiments. This type of data is customarily summarized in dissimilarity/similarity matrices which produce data unfit for ordinary regression analysis because of the dependence structure of the responses. In this chapter two regression approaches, one permutational and the other parametric, are presented to cope with this type data keeping in mind the the objectives of estimation and hypothesis testing.

§ 4.2 *PERMUTATION REGRESSION*

Recent articles by Oja (1987) and Collins (1987) describe the basic approach to permutation simple linear (SLR) and planar regression (PR). Oja (1987) utilizes simplices in construction of test statistics for testing the slope parameter. First, define the SLR model as

$$y = \beta_0 + \beta_1 x + \epsilon \qquad [4.1]$$

Consider $T_1 = \sum_{i<j} \Delta_{ij}^{y} \Delta_{ij}^{x'}$ for testing H_0: $\beta_1 = 0$, where

$$\Delta_{ij}^{y} = y_j - y_i = \begin{vmatrix} 1 & 1 \\ y_i & y_j \end{vmatrix}, \quad i<j, \qquad [4.2]$$

$$\Delta_{ij}^{x'} = x'_j - x'_i = \begin{vmatrix} 1 & 1 \\ x'_i & x'_j \end{vmatrix}, \quad i<j, \qquad [4.3]$$

and x' is a random permutation of $x=(x_1,\ldots, x_n)$.

The model for planar regression is given as

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon \qquad [4.4]$$

where x_2 is assigned at random to experimental units. Testing for H_0: $\beta_2 = 0$, use

$T_2 = \sum_{i<j<k} \Delta^{y}_{ijk} \, \Delta^{x'}_{ijk}$, where

$$\Delta^{y}_{ijk} = \begin{vmatrix} 1 & 1 & 1 \\ y_i & y_j & y_k \\ x_{1i} & x_{1j} & x_{1k} \end{vmatrix} \quad i<j<k, \qquad [4.5]$$

$$\Delta^{x'_2}_{ijk} = \begin{vmatrix} 1 & 1 & 1 \\ x'_{2i} & x'_{2j} & x'_{2k} \\ x_{1i} & x_{1j} & x_{1k} \end{vmatrix} \quad i<j<k, \qquad [4.6]$$

Alternatively, Collins (1987) provides an equivalent statistic for testing the slope which is easier to manipulate. Note that

$$\sum_{i<j} \Delta^{y}_{ij} \, \Delta^{x'}_{ij} = N \sum_{j} y_j x'_j \propto y'x' \qquad [4.7]$$

and

$$\sum_i \sum_j \sum_k \Delta^{y}_{ijk} \, \Delta^{x'}_{ijk} = 6\left\{\left(\sum_i y_i x'_{2i}\right)\left(\sum_i x^2_{1i}\right) - \left(\sum_i y_i x_{1i} x'_{2i}\right)\right\} \propto y'(I-P_1)x'_2 \qquad [4.8]$$

where

$$P_1 = X(X'X)^{-1}X' \text{ and } X = \begin{bmatrix} 1 & x_1 \end{bmatrix}. \qquad [4.9]$$

Defining the test statistic as

$$T_3=\mathbf{e}'\mathbf{x}_2', \tag{4.10}$$

leads to permutation results

$$E_p[T_3|\beta_2=0]=0, \tag{4.11}$$

$$Var_p[T_3]=\frac{(\mathbf{x}_2'\mathbf{x}_2)(\mathbf{e}'\mathbf{e})}{N-1}, \tag{4.12}$$

where $\mathbf{e}=(I-P_1)\mathbf{y}$ is the vector of residuals from regressing $\mathbf{y}$ on $\mathbf{x}_1$. A normal approximation to the permutation distribution of T_3 leads to rejection of H_0: $\beta_2=0$ for

$$|Z|=\mathbf{y}'(I-P_1)\mathbf{x}_2\times\sqrt{\frac{N-1}{(\mathbf{x}_2'\mathbf{x}_2)(\mathbf{e}'\mathbf{e})}} \tag{4.13}$$

exceeding the appropriate critical point of N(0, 1).

A better approximation is presented by Collins (1987) in which the first two permutation moments of T_3 are fitted to a beta distribution. For testing H_0:$\beta_2=0$ in the PR model the F-statistic is

$$F=\frac{(N-2)\mathbf{y}'(P-P_1)\mathbf{y}}{\mathbf{y}'(I-P_1)\mathbf{y}} \tag{4.14}$$

where N is the number of observations, $P=X(X'X)^{-1}X'$, $P_1=X_1(X_1'X_1)^{-1}X_1'$, $X=\begin{bmatrix} 1 & x_1 & x_2 \end{bmatrix}$ and $X_1=\begin{bmatrix} 1 & x_1 \end{bmatrix}$. We can form a beta statistic as

$$U=\frac{y'(P-P_1)y}{y'(I-P_1)y} \qquad [4.15]$$

and if K=1 then,

$$U=\frac{(e'x_2)^2}{(e'e)x_2'(I-P_1)x_2} \qquad [4.16]$$

where $e=(I-P_1)y$ are the residuals of y regressed on x. Under normality of ϵ, $U\sim\beta\left(\frac{1}{2},\ (N-S-1)/2\right)$, where s is the dimension of x_1.

Upon inspection, it becomes apparent that the statistic T_3 can be expressed in the form of a Mantel statistic (eq. 2.26 of § 2.2.5) on the residuals e_{ij},

$$Z=\sum_{i\neq j} X_{2ij}e_{ij}, \qquad [4.17]$$

where $X_{2ij}=X_{2i}-X_{2j}$ and $e_{ij}=e_i-e_j$. Z is the permutation statistic which would be used to test for treatment differences in an ANOVA setting or a partial test of the regression coefficient, $H_0:\beta_2=0$, in a regression situation.

The partial regression coefficient, $\beta_{2,e\cdot x_2}$, associated with the regression of the residuals (from a linear fit of Y on X_1) on X_2 can be calculated as

$$\hat{\beta}_{2,e\cdot x_2}=\frac{\sum_{i\neq j}X_{2ij}e_{ij}-\frac{\sum_{i\neq j}X_{2ij}\sum_{i\neq j}e_{ij}}{n(n-1)}}{\sum_{i\neq j}X_{2ij}-\left[\sum_{i\neq j}X_{2ij}\right]^2\frac{1}{n(n-1)}} \qquad [4.18]$$

$$=\frac{e'x_2-2(e'\underset{\sim}{1})(x_2'\underset{\sim}{1})\binom{n}{2}^{-1}}{x_2'x_2-2(x_2'\underset{\sim}{1})^2\binom{n}{2}^{-1}}. \qquad [4.19]$$

An appropriate test statistic for testing $H_0:\beta_{2,e\cdot x}=0$ and $100(1-\alpha)\%$ confidence interval for $\beta_{2,e\cdot x_2}$ are respectively given as

$$t_{n-2}=\frac{\hat{\beta}_{2,e\cdot x_2}}{S.E.(\hat{\beta}_{2,e\cdot x_2})}, \qquad [4.20]$$

$$\hat{\beta}_{2,e\cdot x_2}\pm t_{1-\frac{\alpha}{2},n-2}S.E.(\hat{\beta}_{2,e\cdot x_2}) \qquad [4.21]$$

where the estimated variance of $\hat{\beta}_{2,e\cdot x_2}$, derived from a method given in Mantel (1967), is given by

$$\mathrm{VAR}(\hat{\beta}_{2,e\cdot x_2})=\left\{\left[\sum_{i\neq j}e_{ij}^2\sum_{i\neq j}X_{2ij}^2-\sum_{i\neq j}e_{ij}e_{ji}\right]+\left\{\left[\sum_i\left(\sum_{j\neq i}e_{ij}^2\right)^2-\sum_{i\neq j}e_{ij}^2\right]\cdot\right.\right.$$

$$\left.\left[\sum_i\left(\sum_{j\neq i}X_{2ij}^2\right)^2-\sum_{i\neq j}X_{2ij}^2\right]+\mathbf{A_eA_x}+2\mathbf{B_eB_x}\right\}/(n-2)+\mathbf{C_eC_x}/(n-2)(n-3)-$$

$$\left.\frac{\left(\sum_{i\neq j}e_{ij}\right)^2\left(\sum_{i\neq j}X_{2ij}\right)^2}{n(n-1)}\right\}\div\left[\sum_{i\neq j}X_{2ij}^2-\frac{\left(\sum_{i\neq j}X_{2ij}\right)^2}{n(n-1)}\right] \qquad [4.22]$$

where,

$$\mathbf{A_e}=\sum_i\left(\sum_{j\neq i}e_{ji}^2\right)^2-\sum_{i\neq j}e_{ji}^2$$

$$\mathbf{A_x}=\sum_i\left(\sum_{j\neq i}X_{2ji}^2\right)^2-\sum_{i\neq j}X_{2ji}^2$$

$$\mathbf{B_e}=\sum_i\left(\sum_{j\neq i}e_{ij}\right)\left(\sum_{j\neq i}e_{ji}\right)-\sum_{i\neq j}e_{ij}e_{ji}$$

$$\mathbf{B_x}=\sum_i\left(\sum_{j\neq i}X_{2ij}\right)\left(\sum_{j\neq i}X_{2ji}\right)-\sum_{i\neq j}X_{2ij}X_{2ji}$$

$$\mathbf{C_e}=\left(\sum_{i\neq j}e_{ij}\right)^2-\mathbf{B_e}-\mathbf{A_e}-\sum_i\left(\sum_{j\neq i}e_{ij}\right)^2-\sum_i\left(\sum_{j\neq i}e_{ij}\right)\left(\sum_{j\neq i}e_{ji}\right)$$

$$\mathbf{C_x}=\left(\sum_{i\neq j}X_{2ij}\right)^2-\mathbf{B_x}-\mathbf{A_x}-\sum_i\left(\sum_{j\neq i}X_{2ij}\right)^2-\sum_i\left(\sum_{j\neq i}X_{2ij}\right)\left(\sum_{j\neq i}X_{2ji}\right)$$

§ 4.3 *REGRESSION MODEL FOR SIMILARITY/DISSIMILARITY DATA*

The regression model presented as equation 2.8 in §2.1.2 is useful in estimating effects for similarity/dissimilarity data as it allows for modelling correlation. In addition to the usual estimation of treatment effects, the regression model will permit us to estimate other important parameters such as the NOEL, which are developed in a parametric framework.

§ 4.3.1 Estimation of Model Parameters

Unbiased estimates for the covariance ρ and variance σ^2 can be obtained by substituting ρ_S in $E(\rho)$ and σ_S^2 in $E(\sigma^2)$ given previously as equations 2.10–2.13. The unbiased estimates, $\hat{\sigma}$ and $\hat{\rho}$, are derived as follows. Substituting σ_S^2 for $E(\sigma_S^2)$ in equation 2.12 we get

$$\sigma_S^2=\sigma^2-\frac{2}{n(n-1)}[(m+1)\sigma^2+tr((X'X)^{-1}X'CX)\rho], \qquad [4.23]$$

similarly we substitute ρ_S for $E(\rho_S^2)$ in equation 2.13,

$$\rho_S=\rho-\frac{\left\{tr[(X'X)^{-1}X'CX]\sigma^2+tr[(X'X)^{-1}X'CX(X'X)^{-1}X'CX]\rho\right\}}{n(n-1)(n-2)} \qquad [4.24]$$

We now have two equations in two unknowns, σ^2 and ρ. Group similar coefficients forming the two equations,

$$\left\{1-\frac{2(m+1)}{n(n-1)}\right\}\sigma^2-2\left\{\frac{tr[(X'X)^{-1}X'CX]}{n(n-1)}\right\}\rho = \sigma_s^2$$

$$\left\{\frac{-tr[(X'X)^{-1}X'CX]}{n(n-1)(n-2)}\right\}\sigma^2+\left\{1-\frac{tr[(X'X)^{-1}X'CX(X'X)^{-1}X'CX]}{n(n-1)(n-2)}\right\}\rho = \rho_s. \quad [4.25]$$

Unbiased estimates for ρ and σ^2 are found by solving for $\hat{\sigma}^2$ and $\hat{\rho}$ in the following linear equations. Let A be defined as

$$\mathbf{A} = \begin{bmatrix} \left(1-\frac{2(m+1)}{n(n-1)}\right) & \frac{-2tr[(X'X)^{-1}X'CX]}{n(n-1)} \\ \frac{-tr[(X'X)^{-1}X'CX]}{n(n-1)(n-2)} & 1-\frac{tr[(X'X)^{-1}X'CX(X'X)^{-1}X'CX]}{n(n-1)(n-2)} \end{bmatrix}$$

then we have

$$\begin{bmatrix} \hat{\sigma}^2 \\ \hat{\rho} \end{bmatrix} = \mathbf{A}^{-1}\begin{bmatrix} \sigma^2{}_s \\ \rho_s \end{bmatrix} \quad [4.26]$$

Solution of the above linear system yields the estimates,

$$\hat{\sigma}^2=\left(\frac{-1}{A_{11}}\right)\sigma_s^2 \quad \text{if } \rho=0, \tag{4.27}$$

$$\hat{\sigma}^2=\frac{\left(A_{22}\sigma_s^2-A_{21}\rho_s\right)}{(A_{11}A_{22}-A_{12}A_{21})} \quad \text{if } \rho\neq 0, \tag{4.28}$$

$$\hat{\rho}=\frac{(-A_{21}\sigma_s^2+A_{11}\rho_s)}{(A_{11}A_{22}-A_{12}A_{21})}, \tag{4.29}$$

where A_{ij} is the ij^{th} entry of A.

If we had substituted the algebraic equivalents of σ_s^2 and ρ_s,

$$\sigma_s^2=\frac{2}{n(n-1)}\left[Y'Y-tr(\hat{\beta}\hat{\beta}'X'X)\right], \tag{4.30}$$

$$\rho_s=\frac{1}{n(n-1)(n-2)}\left[Y'CY-tr(\hat{\beta}\hat{\beta}'X'CX)\right] \tag{4.31}$$

into equations 4.16 and 4.17 and solved for σ^2 and ρ, respectively, we would have obtained,

$$\sigma^2=\frac{2}{n(n-1)-2(m+1)}\left\{Y'Y+tr(\hat{\beta}\hat{\beta}'X'X)+tr((X'X)^{-1}X'CX)\rho\right\} \tag{4.32}$$

$$\rho=\frac{\left\{Y'CY-tr(\hat{\beta}\hat{\beta}'X'CX)+\mathrm{tr}[(X'X)^{-1}X'CX]\sigma^2\right\}}{n(n-1)(n-2)-tr\left\{(X'X)^{-1}X'CX(X'X)^{-1}X'CX\right\}}. \tag{4.33}$$

The following linear system of equations yield solutions for $\hat{\sigma}^2$ and $\hat{\rho}$,

$$\mathbf{B} = \begin{bmatrix} n(n-1)-2(m+1) & -2tr[(X'X)^{-1}X'CX] \\ -tr[(X'X)^{-1}X'CX] & tr[(X'X)^{-1}X'CX(X'X)^{-1}X'CX]-n(n-1)(n-2) \end{bmatrix}$$

$$\begin{bmatrix} \hat{\sigma}^2 \\ \hat{\rho} \end{bmatrix} = \mathbf{B}^{-1} \begin{bmatrix} 2[Y'Y+tr[\hat{\beta}\hat{\beta}'X'X]] \\ Y'CY-tr[\hat{\beta}\hat{\beta}'X'CX] \end{bmatrix} \qquad [4.34]$$

$$\hat{\sigma}^2 = \left(\frac{-1}{B_{11}}\right) 2[Y'Y+tr[\hat{\beta}\hat{\beta}'X'X]] \text{ if } \rho=0, \qquad [4.35]$$

$$\hat{\sigma}^2 = \frac{(2B_{22}\{Y'Y+tr[\hat{\beta}\hat{\beta}'X'X]\}-B_{21}\{Y'CY-tr[\hat{\beta}\hat{\beta}'X'CX]\})}{(B_{11}B_{22}-B_{12}B_{21})} \text{ if } \rho\neq 0, \qquad [4.36]$$

$$\hat{\rho} = \frac{(-2B_{21}\{Y'Y+tr[\hat{\beta}\hat{\beta}'X'X]\}+B_{11}\{Y'CY-tr[\hat{\beta}\hat{\beta}'X'CX]\})}{(B_{11}B_{22}-B_{12}B_{21})}, \qquad [4.37]$$

Dyer gives OLS estimates of the slope parameters as eq. 2.9 from § 2.2.2,

$$\hat{\beta} = (X'X)^{-1}X'Y. \qquad [4.38]$$

OLS estimates may be improved by incorporating a term for the error variance Σ.

Let $VAR(Y)=\Sigma=\sigma^2 I+\rho C$, where $C_{ij}=1$ if Y_i and Y_j have exactly one sample in common. The generalized least squares estimator for β is then

$$\hat{\beta}_{GLS}=(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y \qquad [4.39]$$

with variance–covariance matrix of the coefficient vector given by,

$$VAR(\hat{\beta}_{GLS})=(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}\,VAR(Y)\,\Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}$$

$$=(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}\,\Sigma\Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}=(X'\Sigma^{-1}X)^{-1}. \qquad [4.40]$$

$\hat{\beta}_{GLS}$ may be a better estimator of β when the number of replicates in each treatment group are unequal as well as when variances are nonhomogeneous. OLS and GLS estimates of β and $VAR(\hat{\beta})$ were compared using multivariate data (P=5) generated with $\rho=80$ and $\sigma^2=100$ (Table 4.1). These and other related results indicate when sample sizes are unequal better estimates (i.e., smaller variance) than those suggested by Dyer can be obtained by generalizing the estimators.

Table 4.1.Comparisons of average coefficient estimates and variances obtained from the regression model for dissimilarities (L_2 norm) using 3000 simulations of sample size N=8. Raw data were generated as $MVN_5(\mu_1,\Sigma)$ and $MVN_5(\mu_2,\Sigma)$ where group 1 sample size was $n_1=3$ and group 2 was $n_2=5$. $\mu_1=(100, 100, 100, 100, 100)'$ and $\mu_2=(110, 110, 110, 110, 110)'$.

Regression Method

Least Squares Method	Estimator	Estimate
OLS	$\hat{\beta}_0$	27.825
	$V(\hat{\beta}_0)$	51.648
	$\hat{\beta}_1$	5.566
	$V(\hat{\beta}_1)$	33.555
GLS	$\hat{\beta}_0$	27.891
	$V(\hat{\beta}_0)$	49.068
	$\hat{\beta}_1$	5.444
	$V(\hat{\beta}_1)$	24.565

§ 4.3.2 Decomposition of Regression with Dissimilarity Model

Recall the SLR model with dissimilarity, eq. 2.8, as

$$D_{ij}=\beta_0+\beta_1\delta_{ij}+\epsilon_{ij}$$

The model can be decomposed into two models, one that models only the within treatment dissimilarities and the other modelling the between dissimilarities. Define $B_{ij}=\delta_{ij}$ and $W_{ij}=\delta_{ij}^c$ (i.e., if $\delta_{ij}=1$ then $W_{ij}=0$ and vice versa). Then within dissimilarities can be modelled as

$$D_{w,ij}=\beta_0+\beta_1 W_{ij}+\epsilon_{ij}. \qquad [4.41]$$

The average within dissimilarity between replicates when considering, say, two treatment groups is given as

$$\bar{D}_w=\frac{2\sum_{i<j} W_{ij}D_{ij}}{N_w} \qquad [4.42]$$

$$=2(N_w)^{-1}\sum_{i<j} W_{ij}(\beta_0+\beta_1 W_{ij}+\epsilon_{ij})$$

$$=2(N_w)^{-1}N_w\beta_0+\beta_1(N_w)^{-1}\sum_{i<j} W_{ij}\delta_{ij}+0=\beta_0,$$

where $\sum_{i<j} W_{ij}\delta_{ij}=0$ and $N_w = n_1(n_1-1)+n_2(n_2-1)$. For g groups, $N_w = \binom{N}{2} - \prod_{i=1}^{g} n_i$ $= \sum_{i=1}^{g} n_i(n_i-1)$. The average between dissimilarity between replicates from separate treatment groups is given as

$$\overline{D}_b = \frac{2}{N_b} \sum_{i<j} B_{ij}D_{ij} \qquad [4.43]$$

$$= \frac{1}{N_b} \sum_{i<j} B_{ij}(\beta_0+\beta_1 W_{ij}+\epsilon_{ij})$$

$$= \frac{N_b\beta_0}{N_b} + \frac{\beta_1}{N_b} \sum_{i<j} B_{ij}\delta_{ij} + 0$$

$$= \beta_0+\beta_1$$

where $\sum_{i<j} B_{ij}\delta_{ij}=0$ and $N_b = n_1n_2$. For g groups $N_b = \prod_{i=1}^{g} n_i$.

From equation 4.42 and 4.43 we can interpret β_0 as the average inherent or baseline dissimilarity between treatment groups, and $\beta_0+\beta_1$ as the baseline dissimilarity plus that due to treatment differences. The quantity $\delta = \overline{B}-\overline{W} = \beta_1$ is a measure of the amount of diversity between groups above and beyond baseline dissimilarity. This is the same quantity used in the test statistics for the ANOVA procedure and the MRPP (under conditions specified in § 2.2.4).

A test of $H_0: \beta_1=0$ can be made by comparing the standardized δ to a t distribution with ν degrees of freedom (the resulting standardized delta is the same as

for ANOVA if no correlation is present). A question arises here as to the proper degrees of freedom. Although $n(n-1)/2$ dissimilarity values are used in the regression, there are only n true replicates. As a conservative approach one could use $n-2$ as the degrees of freedom.

If the assumption of normality necessary to employ the t approximation cannot be met, then the test statistic can be tested with a permutation test. In fact, equations 4.42 and 4.43 are already in the form of the Mantel statistic as described in the previous section (§ 4.1.1).

§ 4.3.3 Estimation of D_{75}

Estimation of dose levels for various percentiles of the response variable is common in toxicological microcosm experiments. A primary endpoint is often to estimate the LC_{50} or LD_{50}, the concentration or dose required to kill 50% of the organisms under study. In the present setting, the response is not survival but rather dissimilarity (similarity) between pairs of replicates. We are now interested in the dose required to increase dissimilarity (decrease similarity) by, say, 25%. We shall label this endpoint D_{75}.

The technique of inverse regression as described in Sokal and Rohlf (1969) can be applied to model 4.41 to yield point and interval estimates for D_{75}. Let Y be a vector

of dissimilarities with Y^{75} representing the 75^{th} percentile of the response, then a point estimator for D_{75} is

$$\hat{X}=\frac{Y^{75}-\hat{\beta}_0}{\hat{\beta}_1}, \qquad [4.44]$$

with a $100(1-\alpha)\%$ confidence interval for D_{75} given by

$$\hat{X}+\frac{\hat{\beta}_1(Y^{75}-\overline{Y})\pm t_{\frac{\alpha}{2},N-2}S\sqrt{(\hat{\beta}_1^2-t^2_{\frac{\alpha}{2},N-2}S^2)(1+1/N)+\frac{(Y^{75}-\overline{Y})^2}{\sum_j X_j^2}}}{\hat{\beta}_1^2-t^2_{\frac{\alpha}{2},N-2}S^2} \qquad [4.45]$$

Notice that the interval is symmetric around

$$\hat{X}+\frac{\hat{\beta}_1(Y^{75}-\overline{Y})}{\hat{\beta}_1^2-t^2_{\frac{\alpha}{2},N-2}S^2},$$

not $\hat{X}$.

Narrow limits are achieved by obtaining reliable data, having large sample sizes and by having few dose levels (optimally two) spread as far as possible. There are some circumstances under which this approach is not recommended. There is a large dependency on the model and its validity. If the true model is quadratic, rather than linear, a bias is introduced. A quadratic model may be appropriate for a number of endpoints, such as diversity. To account for nonlinearity, additional doses must be

introduced into the study (Smith and Mercante 1989).

CHAPTER V

CONCLUSIONS AND FURTHER RESEARCH

§ 5.1 *CONCLUDING REMARKS*

Several noteworthy conclusions can be drawn from this dissertation. As a result of the multi–factorial simulation investigation of the power performance of several potentially useful methods, the MRPP and regression with dissimilarity methods emerge as the best methods yet available for the analysis of the complex data arising from microcosm experiments. The MRPP is best suited in situations where replication is present and the response is not normally distributed. Regression with dissimilarities, first proposed by Dyer (1979), provides an additional mechanism for

evaluation of microcosm data. It has the advantage of enabling direct estimation of key parameters not possible with the MRPP and provides a mechanism for quantifying the magnitude of treatment effects.

Variable redundancy was found to have only slight effect on power performance of the MRPP and regression methods when the variables were uncorrelated. A pronounced periodic pattern was evident in the power profile when the data were generated with a correlation of $\rho=0.8$.

A measure of variable importance or influence, initially proposed by Smith (1986), proved useful in dependent variable selection. A stepwise algorithm using a p-value selection criterion was proposed and applied to a real data set. It appears to provide a substantial improvement over the importance measure for selecting important variables.

Permutational and parametric regression methods were developed that provide additional bases for estimation and hypothesis testing for microcosm data. Estimation of D_{75}, a parameter parallel to LC_{50}, was illustrated for dissimilarity data.

§ 5.2 *FURTHER RESEARCH*

Many avenues unexplored or in need of further development remain. Design of multispecies microcosm experiments has taken a traditional and conservative approach. There is much to be done in this area, particularly in exploiting the multivariate nature of the data. Most of the development and results were presented assuming two treatment groups were of interest. Extension to K–groups for all procedures would be a necessary and useful generalization.

Additional research is needed in estimation of key parameters, such as D_{75} and NOEL, so that more informed decisions can be made regarding any deleterious effects a substance may have on the environment. A natural extension of the parametric regression method would be to develop a repeated measures model for microcosm data and derive estimates and tests for model parameters. Exploiting the multivariate nature of the experiments could provide significant improvements over existing methodology in assessing changes in community structure.

BIBLIOGRAPHY

Anderson, T.W. 1958. *Introduction to Multivariate Statistical Analysis.* John Wiley and Sons, New York.

Berry, K.J. 1982. Enumeration of all permutations of multi-sets with fixed repetition numbers. Applied Statistics (31) No. 2 169-173.

Berry, K.J. and P.W. Mielke, Jr. 1984. Computation of exact probability values for multi-response permutation procedures (MRPP). Communications in Statistics, Series B, 13(3):417-432.

Boyle, T.P., J. Sebaugh, and E. Robinson-Wilson. 1984. A Hierarchial approach to the measurement of changes in community structure induced by environmental stress. Journal of Testing and Evaluation 12:241-245.

Brock, D.A. 1977. Comparison of community similarity measures. Journal of Water Pollution Control Federation 49: 2488-2494.

Brockwell, P.J., P.W. Mielke and J. Robinson. 1982. On non-normal invariance principles for multi-response permutation procedures. Austral. J. Statist. 24(1): 33-41.

Brown, C.C. 1978. The statistical analysis of dose-effect relationships. pp 115-148. In: *Principles of Ecotoxicology SCOPE* 12. G.C. Butler, (Ed.) John Wiley and Sons, New York.

Cairns, J., Jr. (Ed.) 1986. *Community Toxicity Testing, STP* 920. American Society for Testing and Materials, Philadelphia.

Cairns, J., Jr. (Ed.) 1985. *Multispecies Toxicity Testing.* Pergammon Press: New York.

Collins, M.F. 1987. A permutation test for planar regression. Austral. J. Statist. 29(3): 303-308.

David, H.A. and B.E. Arens 1959. Optimal spacing in regression analysis. Annals of Math. Stat. 30: 1072-1081.

Dewey, S. L. 1986. Effects of the herbicide atrazine on aquatic insect community structure and emergence. Ecology 67: 148-162.

Dyer, D.P. 1978. An analysis of species dissimilarity using multiple environmental variables. Ecology 59: 117-125.

Foutz, R.V., D.R. Jensen and G.W. Anderson. 1985. Multiple comparisons in the randomization analysis of designed experiments with growth curve responses. Biometrics 41: 29-37.

Giddings, J.M. 1986. Microcosm procedure for determining safe levels of chemical exposure in shallow-water communities, pp 121-134. In: *Community Toxicity Testing*, John Cairns, Jr., (Ed.). American Society for Testing and Materials. Philadelphia.

Giddings, J.M. and G.K. Eddelmon 1979. Some ecological and experimental properties of complex aquatic microcosms. Intern. J. Environmental Studies 13: 119-123.

Giesy, J. P. (Ed.) 1980. *Microcosms in Ecological Research.* CONF-781101, National Technical Information Service, Springfield.

Hruby, T. 1987. Using similarity measures in benthic impact assessments. Environmental Monitoring and Assessment 8: 163-180.

IMSL. 1987. *User's Manual, Stat/Math Library: Fortran Subroutines for Statistical Analysis. Version* 1.0. Houston, TX. pp 1231.

Johnson, R.A. and D.W. Wichern. 1982. *Applied Multivariate Statistical Analysis.* Prentice-Hall: Englewood Cliffs, New Jersey. pp 594.

Kiefer, J. and J. Wolfowitz. 1959. Optimum designs in regression problems. Annals of Math. Stat. 30: 271-294.

Krewski, D. and J. Kovar 1982. Low dose extrapolation under single parameter dose response models. Commun. Statist. - Simula. Computat. 11: 27-45.

Krewski, D., J. Kumar, and M. Bickis. 1984. Optimal experimental designs for low dose extrapolation II. The case of nonzero background. pp 167-191. In: *Topics in Applied Statistics.* Y.P Chaubey and T.D. Dwivedi (eds.). Concordia University, Montreal.

Krewski, D., M. Bickis, J. Kumar and D.L. Arnold 1986. Optimal experimental designs for low dose extrapolation I. The case of zero background. Utilitas Mathematica 29: 245-262.

Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. Cancer Research 27: 209-220.

Mantel, N. and R.S. Valand. 1970. A technique of nonparametric multivariate analysis. Biometrics 26:547-558.

Mielke, P.W. 1986. Non-metric statistical analysis: Some metric alternatives. Journal of Statistical Planning and Inference 13:377-387.

Mielke, P.W. 1979. On the asymptotic non-normality of null distributions of MRPP statistics. Communications in Statistics, Series A, 8(15): 1541-1550.

Mielke, P.W. 1978. Clarification and appropriate inferences for Mantel and Valand's nonparametric multivariate analysis technique. Biometrics (34): 277-282.

Mielke, P.W. 1976. Multi-response permutation procedures. Encyclopedia of Statistics. Johnson and Kotz (Eds.).

Mielke, P.W., K.J. Berry and E. S. Johnson. 1976. Multi-response permutation procedures for a priori classifications. Communications in Statistics, Series A, 5(14): 1409-1424.

Morrison, D.F. 1976. *Multivariate Statistical Methods.* McGraw-Hill, New York, pp 415.

Myers, R.H. 1986. *Classical and Modern Regression with Applications.* Duxbury Press, Boston, 359.

Oja, H. 1987. On permutation tests in multiple regression and analysis of covariance problems. Austral. J. Statist. 29: 91-100.

National Research Council. 1981. *Testing for Effects of Chemicals on Ecosystems.* National Academy Press, Washington.

O'Reilly, F.J. and P.W. Mielke. 1980. Asymptotic normality of MRPP statistics from invariance principles of U statistics. Communications in Statistics, Series A, 9(6): 629-637.

Ricklefs, R.E. and M. Lau. 1980. Bias and dispersion of overlap indices: Results of some Monte Carlo simulations. Ecology 61: 1019-1024.

Robinson, J. 1983. Approximations to some test statistics for permutation tests in a completely randomized design. Austral. J. Statist. 25(2): 358-369.

Seber, G.A.F. 1984. *Multivariate Observations.* John Wiley & Sons, New York. pp 686.

Smith, E.P. 1986. Randomized similarity analysis of multispecies laboratory and field experiments. pp 261-272. In A.H. El-Shaarawi and R.E. Kwiatkowski (eds). *Statistical Aspects of Water Quality Monitoring*, Elsevier, New York.

Smith, E.P. and D.E. Mercante (1989). Statistical concerns in the design and analysis of multispecies microcosm and mesocosm experiments. Toxicity Assessment: An International Journal Vol. 4, 129-147.

Smith, E.P., R.B. Genter, and J. Cairns, Jr. 1986. Confidence intervals for the similarity between algal communities. Hydrobiologia 139: 237-245.

Smith, E.P. and T.M. Zaret. 1982. Bias in estimating niche overlap: Approximate and simulation results. Ecology 63: 1675-1681.

Sokal, R.R. and F.J. Rohlf. 1969. *Biometry.* pp 776. W. H. Freeman and Co. San Francisco.

Stander, J.M. 1970. Diversity and similarity of benthic fauna off the coast of Oregon. M.S. Thesis Oregon State University, Corvallis, Oregon. 72 pp.

Taub, F.B. 1976. Demonstration of pollution effects in aquatic microcosms. Intern. J. Environmental Studies, Vol. 10: 23-33.

Tracy, D.S. and I.H. Tajuddin. 1986. Emperical power comparisons of two MRPP rank tests. Communications in Statistics, Series A, 15(2): 551-570.

Tracy, D.S. and I.H. Tajuddin. 1985. Extended moment results for improving inferences based on MRPP. Communications in Statistics, Series A, 14(6) 1485-1496.

VITA

Donald Eugene Mercante was born on July 18, 1955 in New Orleans, Louisiana. After attending parochial elementary and high schools in New Orleans he enrolled in Nicholls State University, receiving his B.S. degree in Marine Biology in 1977. He and Teresita Musacchia of New Orleans were married on July 28, 1978. He began graduate studies at Mississippi State University in August, 1978, completing his M.S. degree in Fisheries Management in August, 1980.

Pursuing an interest in statistics cultivated while at MSU, the author enrolled in the Department of Experimental Statistics at Louisiana State University. He was awarded the Master of Applied Statistics degree from LSU in 1983. He was employed as a biostatistician in the Department of Pathology at LSU Medical Center from 1983–1985 working under Professor Miguel Guzman. He made the decision to further his education and in the Fall of 1985 enrolled as a Ph.D. student in the Department of Statistics at Virginia Polytechnic Institute and State University.

The author accepted a position as Senior Research Biostatistician with the Merrell Dow Research Institute of Cincinnati, Ohio in February, 1989, where he now lives with his wife, Teresita, and their two daughters, Lorenza Maria and Anna Alicia.