Some Views on Parametric and Non-Parametric Analysis for Repeated Measurements and Selected Bibliography

Gary G. Koch, Ingrid A. Amara, Maura E. Stokes, and Dennis B. Gillings

Department of Biostatistics, University of North Carolina, Chapel Hill, N.C.

Summary

A common feature of many statistical investigations is the collection of data from groups of experimental units each of which is observed under two or more conditions. Such studies are generally called either split-plot experiments or repeated measurements experiments. This paper is concerned with reviewing general statistical strategies for the analysis of data from these types of research designs. For this purpose, primary attention is directed at two basic dimensions. One of these is the nature of the randomization processes for the data as obtained among and within the experimental units. The other is the level of the measurement scale as either nominal, ordinal, or interval. This framework is then used as the basis of discussion of alternative statistical methods such as repeated measurements analysis of variance, multivariate analysis of variance, and their non-parametric rank and categorical data counterparts in both a general sense and for some specific classes of examples. Finally, a selected bibliography of references for these and related methods is given.

1 Introduction

The following discussion is a survey of methodological issues pertaining to the statistical analysis of investigations involving repeated measurements. Although such research designs arise in many different ways, the principal consideration which distinguishes them from standard factorial experiments is the study of a randomly selected or allocated experimental unit under two or more distinct conditions (*e.g.*, repeated measurements over time) which represent observational units. Thus, the linkage between experimental units and observational units is perhaps the most important aspect of the formulation of relevant statistical methods. Against this background, the experimental units can be viewed as the natural units for analysis and their observational units can be viewed as a profile of inherently multivariate data. Thus, analyses can be formulated in terms of operations whereby certain measures of interest (*e.g.*, sums and differences, orthogonal contrasts, regression coefficients, *etc.*) are first constructed from the observational unit data within each experimental unit and then summarized across experimental units in either a parallel, but separate, univariate sense and/or in a simultaneous multivariate sense. Since this approach makes no assumptions concerning the covariance structure of the observational units, it is reasonable regardless of whether the observational units correspond to randomized or non-randomized conditions. On the other hand, when the observational units have randomly assigned conditions (and so are secondary sampling units), then they may often be analyzed more effectively on a within experimental unit basis because of compound symmetry properties (*i.e.*, equal variances and equal covariances). For this reason, the nature of randomization processes for investigations with repeated measurements must always be taken into account because of their implications to the linkage of experimental and observational units. In the remainder of this paper, several different types of situations which involve repeated measurements are described. Then available methods for the statistical analysis of repeated measurements are summarized in a way that expresses their relationship to the randomization process for the research design and the level of measurement for the data (categorical, ordinal, interval). Finally, specific recommendations concerning analysis are provided for some of the more common types of applications.

2 Some examples of repeated measurement investigations

The following examples are illustrative of the range of application of repeated measurement studies in the biological, social and physical sciences

1. An agricultural experiment is concerned with the comparison of several fertilizers and several methods of cultivation of some crop. For this purpose, the fertilizers are first randomly assigned to the fields which are to be used. Each field is then divided into plots to which the methods of cultivation are randomly assigned. Since method of cultivation to plots is randomly assigned within the random assignment of fields to fertilizers, such experiments are often called split-plot experiments. Here, fields represent the whole plots and plots represent the split plots. Alternatively, fields represent the experimental units to which the primary or first level randomization is applied, and plots represent the observational units to which the second level of randomization is applied. Thus, such experiments involve two levels of randomization and this aspect of their structure is of fundamental importance to their analysis.
2. Mother animals are assigned to different diet treatments during pregnancy. Then the baby animals in their litters are randomly assigned to different growth diet conditions. The data under study are the weights of the baby animals at the end of each of 10 successive weeks (where the initial week begins at 6 weeks of age). Here, mothers represent the experimental units (whole plots) to which the primary or first level of randomization is applied, and babies represent the observational units to which the second level of randomization is applied. Moreover, week represents a third level non-randomized observational unit.
3. An industrial experiment is concerned with the effects of five (two-level) control factors on the performance of an instrument for assessing the processability of a polymer in terms of the torque required to stir the polymer as a function of time. Runs are undertaken for the $2^5 = 32$ combinations of these factors in a random order, and data for the torque values at 4, 5, 7.5, 10, 12.5, and 15 minutes are obtained from the resulting response curves for purposes of analysis. Here, runs (or response curves) represent randomly allocated experimental units and time corresponds to non-randomized observational units (*i.e.*, conditions).
4. Bacteria samples are obtained from the throats of male and female children in the 4–12 year age range. From each sample, three cultures are prepared and are randomly assigned to three drugs for sensitivity testing. Here, children represent a non-randomized experimental unit study population, and cultures represent a randomized observational unit.
5. Dairy cows are randomly assigned to several methods of treatment for mastitis. Data pertaining to the disease status of each quarter of the udder of each cow are obtained both before and after treatment. Here, cows represent a randomized experimental unit, and udder quarter and time represent non-randomized observational units (*i.e.*, conditions).
6. A multi-clinic clinical trial is undertaken to compare two treatments. For this purpose, a specific set of appropriately qualified clinics are invited to participate on a judgmental basis. Within each clinic, patients are randomly assigned to the two treatments. Data pertaining to various aspects of the medical status of each patient are recorded both before treatment (*e.g.*, baseline) and at weekly visits during treatment for a 6 week period. Here, clinics represent a non-randomized set of study sub-populations; patients represent a randomized experimental unit; and visits represent a non-randomized observational unit

7. An agricultural study is undertaken to investigate the effects of variety and planting date on the size and shape of sweet potatoes. Data pertaining to the length and diameters at four positions are obtained from twenty samples at each of six homogeneous plots for a three varieties × 2 planting dates factorial design. The shape of each of the sweet potatoes is then summarized in terms of diameter/length ratios for the four measurement positions. Here, plots are viewed as a fixed set of homogeneous study populations to which the six varieties × planting dates combinations may or may not necessarily have been randomly allocated. Thus, sweet potatoes represent within plot randomly selected experimental units and positions for the diameter/length ratios represent non-randomized observational units.
8. A multi-period change-over clinical trial is undertaken to compare two treatments. For this purpose, patients are randomly partitioned into sequence groups which define the order according to which they are given the two treatments during successive periods of time. For example, some patients receive the control treatment first followed by the new treatment while others receive the new treatment first followed by the control treatment. However, more than two sequence groups and/or more than two periods can be used. Finally, data pertaining to the health status of each patient are obtained at weekly visits for each of the 4 week treatment periods over which treatments are applied, at pre-treatment, and at the end of a no-treatment washout week between the two 4 week treatment periods. Here, patients represent a randomized experimental unit with respect to sequence group; and period represents a non-randomized observational unit. Furthermore, weekly visits represent third level non-randomized observational units.
9. A political survey is conducted to compare the level of trust which a sample of subjects have for three political institutions – the Presidency, the Supreme Court, and the Senate. Data pertaining to attitudes toward the three institutions are obtained for each subject. Here, subjects represent a randomly selected group of primary sampling units (*i.e.*, experimental units) and attitudes represent non-randomized observational units (*i.e.*, conditions).
10. A behavioral development study is conducted longitudinally to compare boys and girls from a 1-year-old age cohort with respect to ability to perform a particular task at the age of 1 year, 2 years, and 4 years. Here, children represent a randomly selected group of primary sampling units (experimental units) and time represents a set of non-randomized observational units.
11. A political survey is conducted to compare the preferences of a sample of subjects to seven tax-expenditure policy alternatives. For this purpose, each subject is asked to rank the seven alternatives from 'most desirable' to 'least desirable'. Here, subjects represent a randomly selected group of primary sampling units (*i.e.*, experimental units) and the tax alternatives represent non-randomized observational units.
12. An observer agreement study is conducted with respect to some phenomena like blood pressure measurement or disease diagnosis. For this purpose, data pertaining to some health characteristic are obtained from each of several observers for all patients in a sample from the population of interest. Here, patients represent a randomly selected group of experimental units, and observer determinations represent non-randomized observational units. (*i.e.*, conditions).

3 Statistical methods for repeated measurements investigations

For the various types of examples described in (1)–(12) of Section 2 the basic data array under consideration can usually be summarized as shown in Table 1. There $h = 1, 2, \ldots, q$ indexes a set of subpopulations or strata (*e.g.*, clinics) within which stratified random selection or random assignment is undertaken; $i = 1, 2, \ldots, s$ indexes a set of groups of experimental units which are to be compared; $j = 1, 2 \ldots, d$ indexes a set of conditions for which observational unit responses are obtained; $k = 1, 2, \ldots, t$ indexes a set of covariables which are determined for each experimental unit as a whole; and $m = 1, 2, \ldots, n_{hi}$ indexes the set of experimental units in the ith group for the hth stratum. Finally, y_{hijm} denotes the response for the jth condition observational unit within the mth experimental unit of the ith group for the hth stratum and x_{hikm} denotes the value of the kth covariable for the mth experimental unit of the ith group for the hth stratum. Thus, for the situation in (6) of Section 2, $h = 1, 2, \ldots, q$ indexes clinics; $i = 1, 2$, indexes treatments; $j = 1, 2, \ldots, 6$ indexes weekly visits; $k = 1, 2, 3$ indexes patient (*i.e.*, experimental unit) covariables such as age (x_{hi1m}), sex (x_{hi2m}), baseline status (x_{hi3m}); $m = 1, 2,$

Table 1

Observed Data Array for h-th Stratum

Sub-population (group)	Experimental unit (subject) in group	Conditions: Response Variables (observational units) 1	2	...	d	Covariables 1	2	...	t
1	1	y_{h111}	y_{h121}	...	y_{h1d1}	x_{h111}	x_{h121}	...	x_{h1t1}
1	2	y_{h112}	y_{h122}	...	y_{h1d2}	x_{h112}	x_{h122}	...	x_{h1t2}
...	...	...	...	...	...	...	...	...	...
1	n_{h1}	$y_{h11n_{h1}}$	$y_{h12n_{h1}}$	...	$y_{h1dn_{h1}}$	$x_{h11n_{h1}}$	$x_{h12n_{h1}}$	...	$x_{h1tn_{h1}}$
1	Mean	$\bar{y}_{h11\cdot}$	$\bar{y}_{h12\cdot}$	...	$\bar{y}_{h1d\cdot}$	$\bar{x}_{h11\cdot}$	$\bar{x}_{h12\cdot}$	...	$\bar{x}_{h1t\cdot}$
2	1	y_{h211}	y_{h221}	...	y_{h2d1}	x_{h211}	x_{h221}	...	x_{h2t1}
2	2	y_{h212}	y_{h222}	...	y_{h2d2}	x_{h212}	x_{h222}	...	x_{h2t2}
...	...	...	...	...	...	...	...	...	...
2	n_{h2}	$y_{h21n_{h2}}$	$y_{h22n_{h2}}$	...	$y_{h2dn_{h2}}$	$x_{h21n_{h2}}$	$x_{h22n_{h2}}$	...	$x_{h2tn_{h2}}$
2	Mean	$\bar{y}_{h21\cdot}$	$\bar{y}_{h22\cdot}$	...	$\bar{y}_{h2d\cdot}$	$\bar{x}_{h21\cdot}$	$\bar{x}_{h22\cdot}$	...	$\bar{x}_{h2t\cdot}$
...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...
s	1	y_{hs11}	y_{hs21}	...	y_{hsd1}	x_{hs11}	x_{hs21}	...	x_{hst1}
s	2	y_{hs12}	y_{hs22}	...	y_{hsd2}	x_{hs12}	x_{hs22}	...	x_{hst2}
...	...	...	...	...	...	...	...	...	...
s	n_{hs}	$y_{hs1n_{hs}}$	$y_{hs2n_{hs}}$	...	$y_{hsdn_{hs}}$	$x_{hs1n_{hs}}$	$x_{hs2n_{hs}}$	...	$x_{hstn_{hs}}$
s	Mean	$\bar{y}_{hs1\cdot}$	$\bar{y}_{hs2\cdot}$	...	$\bar{y}_{hsd\cdot}$	$\bar{x}_{hs1\cdot}$	$\bar{x}_{hs2\cdot}$	...	$\bar{x}_{hst\cdot}$
Overall	Mean	$\bar{y}_{h\cdot1\cdot}$	$\bar{y}_{h\cdot2\cdot}$	...	$\bar{y}_{h\cdot d\cdot}$	$\bar{x}_{h\cdot1\cdot}$	$\bar{x}_{h\cdot2\cdot}$	...	$\bar{x}_{h\cdot t\cdot}$

$\ldots, n_{hi}$ indexes patients; and y_{hijm} represents some measure of medical status which is determined at jth week for mth patient with ith treatment in hth clinic. Otherwise, if there are several measures of medical status, the y_{hijm} are generalized to a corresponding vector; similarly, such extensions can also encompass covariables pertaining to the observational units.

3.1 *Repeated measurements (split-plot) analysis of variance*

This approach is based on the partition of the pertinent sources of variation into two or more basic types. One of these is the variation among experimental units with respect to averages (or sums) of the observational units, and the other is the variation among observational units within experimental units. Since the sampling error asssociated with variation among experimental units is usually larger than the response (or sub-sampling or measurement) error for observational units within experimental units, these types of variation are addressed separately (even though their results are usually presented together). In this regard, they each involve components which are directed at certain hypotheses of interest and corresponding error variance measures.

The most well-known framework for repeated measurement analysis of variance is directed at *intervally scaled quantitative data* which can be assumed to have normal distributions. Furthermore, homogeneous variance for observational units is assumed (both within and across experimental units); homogeneous covariance for pairs of observational units within the same experimental unit is assumed (both within and across experimental units); and independence of experimental units is assumed. Given these specifications together with a balanced experiment with no missing data, it then follows that hypothesis testing can be undertaken with *F*-statistics (provided that the experiment is not complicated by other difficulties, such as aliasing and/or confounding of within experimental unit factors; *e.g.*, periods and treatments in multi-period change-over designs).

More generally, certain aspects of repeated measurement analysis of variance can be justified by randomization arguments which are essentially the same as those underlying analogous non-parametric methods. In this context, the corresponding *F* statistics are approximate if the sample size is sufficiently large with respect to the number of experimental units (*e.g.*, at least 20 per treatment group) for central limit theory to be applicable. Thus, assumptions concerning normality, homogeneous variance, and homogeneous within experimental unit covariance are not strictly necessary. On the other hand, the nature of the randomization process is critical. In particular, randomization of experimental units to treatment groups is required for the validity of the *F*-statistic for comparing treatment groups (with respect to observational unit sums for experimental units); and randomization of conditions to observational units is required for the validity of the *F*-statistic for comparing conditions (within experimental units). However, the *F*-statistic for treatment × condition interaction cannot be justified by randomization arguments since treatment group differences and condition differences contradict the equal probability structure for permutations of experimental units among treatment groups and observational units among conditions. Thus, if this source of variation is of direct interest, then it must be investigated by some other procedure like the multivariate methods discussed in Section 3.2.

Finally, if the sample size under study is not large and the assumptions concerning normality and/or covariance structure are considered unrealistic, then the significance of repeated measurement analysis of variance test statistics should be based on exact permutation distributions as opposed to *F* approximations. However, in these situations, non-parametric rank methods may be more convenient because tables of their exact distributions are already available to a broad extent. Also, since rank methods involve uniform distributions without extreme values, their central limit theory approximations may be more reasonable for moderate sample sizes than the interval scale framework with respect to which the data are actually observed.

In summary, if the data under study are intervally scaled, then sums and differences with respect to them have meaningful definitions. When randomization arguments are applicable, these quantitative measures can be analyzed by exact permutation methods or large sample *F* tests in a non-parametric framework which does not require assumptions concerning the underlying distribution of the data. Thus, the structure of such test statistics as it relates to the hypotheses under investigation can be viewed as parametric even though the procedure by which statistical significance is evaluated is non-parametric. Of course, if the data do satisfy the necessary assumptions of normality, homogeneous variance, and homogeneous covariance, then both of these considerations are parametric. Alternatively, if rank methods are used, both are non-parametric. As a result, several types of parametric and/or non-parametric strategies are potentially applicable to intervally scaled data depending upon the nature of the randomization process, the sample size available, and the plausibility of assumptions concerning their distributions.

3.2 *Repeated measurement multivariate profile analysis of variance*

Since the set of conditions for the observational units within each experimental unit can be viewed as a multivariate set of responses, multivariate statistical methods represent the most natural framework for the analysis of repeated measurements. With this approach, partition of variation into pertinent sources is only undertaken with respect to the sampling error associated with experimental units. However, this strategy is applied to multivariate sets of linear functions of observational units, each of which has its corresponding error covariance structure. Thus conditions within experimental units can be compared through the use of difference functions (or other types of contrasts) in a manner which does not require that any direct consideration be given to variation among the corresponding observational units. However, the nature of such observational unit variation is taken into account indirectly through the estimated covariance structure which is used with such multivariate methods. Similarly, treatment groups can be compared with respect to experimental unit averages over conditions through the use of sums (or means) for the corresponding observational units. Thus, with the appropriate choice of functions, a broad range of hypotheses can be investigated via multivariate strategies.

When intervally scaled quantitative data are under consideration, the usual assumptions for justifying multivariate analysis of variance methods are that the vectors of observational units for the respective experimental units have independent multivariate normal distributions and that they have the same covariance matrix (which necessarily implies that they have the same dimension and hence that there are no missing data). Under these conditions, the characteristic roots of the products of hypothesis sums of product matrices post-multiplied by the inverse of the error sums of product matrices can be used to test various hypotheses. For small samples, special tables are needed to evaluate the significance of such tests for all situations except the special cases where F-distributions are applicable (*i.e.*, single univariate function analyses of variance, Hotelling T^2 for testing whether a single mean vector equals a known constant vector or testing the equality of two mean vectors, *etc.*). Thus, the usefulness of multivarate methods is sometimes limited by the extra effort which is required for interpreting their results in such cases. However, if the sample size is sufficiently large with respect to the number of experimental units (*e.g.*, at least 30 per treatment group) and the number of observational units per experimental unit is small (*e.g.*, not greater than five), chi-square approximations are potentially applicable to multivariate trace criteria on the basis of central limit theory. Furthermore, the use of chi-square approximations can also be justified by randomization arguments with respect to the process according to which experimental units are assigned to groups (see Koch and Bhapkar (1980) or Puri and Sen (1971) for further discussion). Thus, the assumptions concerning multivariate normality and homogeneous covariance structure are not strictly necessary. With this consideration in mind, the principal advantage of the multivariate analysis approach over the split-plot approach is that it permits valid tests to be undertaken for within experimental unit hypotheses concerning the comparison of conditions for situations where there is no randomization of conditions to observational unit. Similarly, it also permits tests to be undertaken for condition × treatment interaction. On the other hand, it may either be inapplicable or have unsatisfactory power when the number of observational units per experimental unit either exceeds or is relatively large with respect to the number of experimental units. Thus, for these types of situations, the scope of analysis must be reduced to parallel analyses of small sets of linear functions of observational units which are summary measures of the variation among conditions with respect to specific research questions of interest (*e.g.*, weekly averages of extent of adverse experiences or concomitant medication usage for daily diary data in clinical trials; growth curve parameters in animal nutrition experiments, *etc.*). Otherwise, within experimental units randomization of conditions to observational unit arguments must be applicable in order to justify the use of split-plot methods to investigate the variation among conditions in an overall sense (as opposed to a summary measure sense). Although the issues presented here concerning the multivariate approach vs the split-plot approach appear somewhat complicated, their implications to most practical situations are straightforward. In particular, when conditions correspond to certain additional aspects of treatment which are of direct interest in their own right, then randomization of conditions to observational units is usually undertaken so that the split-plot approach is valid and also effective with respect to power. For this reason, the multivariate approach would only be needed in order to investigate treatment × condition interaction which could be limited to summary measures which the split-plot approach identified as the most important sources of variation among conditions within experimental units. Alternatively, when conditions correspond to dimensions which reflect multivariate aspects of the response of experimental units (*e.g.*, the responses at different points in time for longitudinal studies, the responses at different sites for certain body system studies, the responses to different questionnaire items for certain psychological or social sciences studies, *etc.*), then the multivariate approach is considered preferable because its validity is linked only to among experimental unit randomization. Moreover, since the conditions often represent components of a profile in such situations, their reduction to summary indexes is usually compatible with both the statistical and substantive objectives for their analysis. Otherwise, such investigations also often involve relatively large sample sizes, and so the multivariate approach can be applied in a straightforward manner via chi-square approximations and be effective with respect to power (*i.e.*, the multivariate approach is asymptotically equivalent to the split-plot approach when the split-plot assumptions are true because their corresponding estimated covariance structures have the same limit; it is also asymptotically equivalent to approaches based on other types of covariance structure assumptions when they are correspondingly applicable).

Finally, for small sample size situations where multivariate normality and homogeneous covariance structure assumptions are considered unrealistic, the significance of multivariate analysis of variance test statistics should be based on exact permutation distributions as opposed to multivariate criterion small sample exact tables. Alternatively, chi-square approximations for multivariate rank methods could be used if the sample size is moderate (*e.g.*, at least 30 experimental units per treatment group) since their central limit theory basis may be more realistic than their counterparts with respect to the interval scale according to which data are observed. Thus, several types of parametric and/or non-parametric strategies, as summarized in Section 3.2, are potentially applicable to the multivariate framework discussed here.

3.3 *Non-parametric rank methods*

3.3.1 *Rank analysis of variance methods* Kruskal-Wallis rank analysis of variance methods can be used to investigate among experimental unit sources of variation. For this purpose, attention is directed at univariate or multivariate functional scores which are constructed from the data for their observational units in accordance with specifications implied by the hypotheses of interest. Thus, treatment group comparisons can be undertaken with respect to questions concerning

1. each condition separately via the corresponding set of ranks for the pooled set of experimental units;
2. all conditions simultaneously via the corresponding array of ranks for the pooled set of experimental units;
3. each condition separately or all conditions simultaneously with covariance adjustment via the corresponding array of rank residuals for the pooled set of experimental units (see Quade (1967));
4. condition averages or sub-averages via ranks for such averages or averages for ranks for the pooled set of experimental units (see Koch (1969, 1970, 1972));
5. summary condition difference scores via difference scores for ranks or summary scores for ranks of differences for the pooled set of experimental units (see Koch (1969, 1970, 1972));
6. measures of trend or change via ranks for such measures for the pooled set of experimental units;

7. other rank score functions for the pooled set of experimental units like those used to analyze censored survival data (*i.e.*, combined scores derived from paired comparisons of experimental units).

For all of these situations, the only necessary assumption is that randomization arguments be applicable with respect to the assignment of experimental units to treatment groups. Also those strategies which involve ranks for averages, differences, or other linear functions require that the data have been observed on an interval scale in order for the corresponding transformed values to be meaningfully defined. Finally, an important specification for purposes of interpretation is that measurement error be negligible for the response variables under study or that the measurement error process be independent of the randomization process. Here, the principal concern is that data quality should not be confounded with the structure of the treatment groups and hence not be a contaminant to their comparison. Given this background, the types of non-parametric rank methods described in (1)–(7) represent valid and useful strategies for the analysis of randomized experiments on the basis of their designs. They may also be applied in other contexts, such as:

i. experiments with a randomization framework which is less restricted than the one actually used in the research design (*e.g.*, randomization of successive subsets of patients is often viewed as equivalent to randomization of the entire set);
ii. observational data for which the concept of randomization is incorporated within the hypothesis (see Koch, Gillings, and Stokes (1980)).

However, more caution is needed with respect to the interpretation of results for these latter situations (since any differences which are detected among the treatment groups may be a consequence of the relationship of both the response variable and the groups to some other variable which has not been taken into account). Otherwise, for moderately large samples (*i.e.*, at least 10 and preferably at least 20 experimental units per treatment group) statistical significance of rank analysis of variance can be assessed via chi-square approximations while for small samples, exact permutation distributions should be used.

3.3.2 *Partial association rank methods* If conditions are randomly assigned to observational units, then partial association rank methods can be used for comparisons within experimental units. Here, the most well-known procedure is Friedman's chi-square statistic which is based on the direct internal ranking of observational units within experimental units. Then the ranks for the conditions are added across experimental units together with their corresponding randomization expected values and covariance matrices; and from these quantities, test statistics with large sample chi-square distributions can be computed. Alternatively, this same type of strategy can be applied to other types of rank functions. In this regard, if the data are observed on an interval scale, then comparisons among conditions can be undertaken via aligned rank tests (see Koch and Sen (1968)). These methods are based on overall ranks for differences between measurements for each observational unit and its experimental unit mean. Otherwise, if the number of experimental units is large (*e.g.*, at least 10 and preferably at least 20), the statistical significance of these partial association rank statistics can be assessed via chi-square approximations while if it is small, exact permutation methods should be used.

3.3.3 *Signed rank methods* For those situations where conditions are not randomly assigned to observational units, signed rank methods can be used for comparisons within experimental units. In this regard, there are two basic strategies. One of these is a within experimental unit rank-oriented generalized sign test which is applicable to both ordinal and interval data. Thus, it is based on the same rank matrix as the Friedman chi-square statistics discussed in Section 3.2.2. However, here the hypothesis of interest is not based on randomization principles but rather on inversion principles which imply that an observed outcome and its opposite are equally likely for each experimental unit. From this framework, expected values and covariance matrices are then determined for the ranks within each experimental unit. Then these quantities are all added across experimental units and combined together to form test statistics with large sample chi-square distributions (see Koch and Sen (1968)). Alternatively, if the data are observed on an interval scale, then a Wilcoxon signed rank approach can be used. For this purpose, pairwise differences between conditions are formed within each experimental unit and assigned signed ranks across experimental units. Then within experimental unit scores are computed for each condition by combining together the signed ranks for the pairwise differences between it and all other conditions. After this score matrix is obtained, inversion principles are used to obtain a generalized sign rank test statistic in the same manner described for the generalized sign test. Otherwise, if the number of experimental units is large (*e.g.*, at least 20 and preferably at least 30), the statistical significance of these signed rank statistics can be assessed via chi-square approximations while if it is small, exact permutation methods should be used (see Koch and Sen (1968) and Koch (1969, 1970)).

3.3.4 *Descriptive statistics* Although the rank methods discussed here are useful for assessing the statistical significance of sources of variation such as groups, conditions, and group × condition interaction, they do not provide direct information about the nature or magnitude of the corresponding differences under investigation. For this reason, non-parametric statistical tests should be accompanied by appropriate descriptive statistics such as frequency distributions or percentiles. Also, when applicable, non-parametric procedures for estimation and confidence intervals represent another method of interest (see Conover (1971) and Puri and Sen (1971)).

3.3.5 *Summary statement* Rank analysis of variance methods in combination with signed rank methods represent the non-parametric analogues for the multivariate approach described in Section 3.2. As such, they are based on among experimental unit randomization and within experimental unit inversion concepts. Partial association rank methods involve within experimental unit randomization and hence are analogous to corresponding methods for the split-plot approach. Thus, essentially the same components concerning the relative advantages of the multivariate approach vs the split-plot approach could be stated for their non-parametric counterparts.

3.4 *Categorical data methods*

3.4.1 *Multivariate non-parametric procedures* All of the non-parametric strategies described in Section 3.3 are directly applicable to categorical data in accordance with their corresponding randomization or inversion principles. In this regard, if the categorical data are ordinally scaled, then rank analysis of variance methods, partial association rank methods, and signed rank methods can be used directly without any special modifications. In other words, the only difference between this type of data and other types of ordinal data is the potentially extensive prevalence of ties which need to be taken into account via midranks. However, two peripheral issues should be recognized here. The first is that interval scale oriented non-parametric methods should not be applied to ordinal categorical data unless they make sense (*e.g.*, they might be appropriate for number of infected sites or size of infected area but not appropriate for subjective ratings of extent of improvement). The other consideration is that larger sample sizes are needed for the use of chi-square approximations for ordinal categorical data than quantitative measurement data since the presence of extensive ties potentially may decrease the speed of convergence for central limit theory (see Koch and Bhapkar (1980)).

Alternatively, if the categorical data are nominally scaled, then they are first transformed to a multivariate vector of binary indicator variables with dimension equal to $(r - 1)$ where r denotes the number of nominal categories. Then multivariate extensions of non-parametric rank methods are applied directly to these vectors with ties being taken into account via midranks. Moreover, interval scale oriented methods are of interest here since a binary scale has this structure in a degenerate sense. Finally, since these methods potentially may involve large numbers of degrees of freedom for strongly discrete variables, very large sample sizes may be required if chi-square approximations are to be used to assess statistical significance.

3.4.2 *Multivariate linear model procedures* If the number of experimental units in each treatment group is moderately large (*e.g.*, at least 20, preferably at least 40 and if rare events are involved, perhaps at least 100), then linear or log-linear functions of summary measures of their experience with respect to categorical variables for the respective conditions can often be presumed to have a multivariate normal distribution. In addition, consistent estimates of their covariance structure usually can be determined by linear Taylor series matrix methods and can be viewed for purposes of estimation and hypothesis testing as essentially the same as the actual covariance structure. Given these considerations, weighted least squares asymptotic regression methods can be used to investigate both among and within experimental unit sources of variation in a manner which is directly analogous to the multivariate approach described in Section 3.2. Furthermore, this same type of strategy is applicable to quantitative data as well as categorical data if the functions of interest are within treatment group arithmetic means or log geometric means; and thus it represents a generalized multivariate approach for dealing with non-normality and/or heterogeneous across group covariance structure issues. In other words, it is based on the framework for which experimental units in the respective treatment groups are regarded as a stratified random sample from a similarly structured hypothetical super-population rather than on the framework for which a finite population randomization process is used to assign experimental units to treatment groups. Thus, such methods are of interest for investigating various aspects of the relationships between response variables and other characteristics of experimental units in addition to treatment group for the analytical situation which potentially exists when treatment groups are actually found to be different, and hence there is specific interest in the nature of such differences. Here, the fundamental issue is that randomization procedures can only be applied to investigate whether hypotheses are contradicted or not. For this purpose, they are very useful because they require essentially no assumptions beyond the structure which is built into the research design. However, once a hypothesis is rejected, randomization concepts are no longer applicable because the randomization probability structure is no longer valid. Thus, other methods must be used to investigate the nature of differences which are detected. However, the formulation of these methods must be undertaken with caution because their assumptions can no longer be reinforced by randomization principles. Given this background, multivariate linear model methods represent one potentially useful strategy which has the advantage of essentially no assumptions other than random sampling, but the disadvantage of large sample size requirements. Alternatively, maximum likelihood methods analogous to multivariate analysis of variance are also of interest if the assumptions for their underlying probability models are considered realistic since the sample size requirements for their central limit theory may not be as stringent as for their assumption-free weighted least squares counterparts (see Koch, Gillings, and Stokes (1980)).

3.4.3 *Summary statement* One way of dealing with categorical data involving repeated measurements is to use non-parametric rank methods with ties handled via mid-ranks. Here, some commonly known linkages are as follows:

i. randomization model ridit analysis is the same as rank analysis of variance;
ii. Cochran's Q test is the same as Friedman's chi-square test;

iii. Mantel-Haenszel tests for marginal homogeneity, symmetry and independence are the same as multivariate Friedman chi-square tests;

iv. Fisher's exact test for treatment effects with respect to a binary response in the two-period change-over design is the same as rank analysis of variance for the subset of non-null differences between the responses for the two periods.

Alternatively, for situations where sample sizes are large and/or certain assumptions concerning distributional structure are realistic, multivariate linear model methods are of potential interest for investigating the effects of treatments groups and other factors on response variables. Here, two useful approaches are maximum likelihood and weighted least squares.

3.5 *Relationships among alternative methods*

The relationships among the alternative methods for the analysis of repeated measurements are summarized in Table 2 with respect to measurement scale and randomization structure.

Table 2

Randomization Structure	Interval Data	Ordinal Data	Nominal Data
Randomization of experimental units to treatment groups	Repeated measurement multivariate analysis of variance	Rank analysis of variance methods (Kruskal-Wallis procedures) and signed rank methods (generalized sign procedure and Wilcoxon signed rank procedure)	Multivariate chi-square tests for homogeneity
Randomization of observational units to conditions	Repeated measurement (split-plot) analysis of variance	Partial association rank methods (Friedman chi-square and aligned rank procedures)	Generalized Mantel-Haenszel methods and Cochran Q-statistic
Stratified random selection of experimental units from possibly different populations	Repeated measurement multivariate analysis of variance with assumed distribution and covariance structure	Multivariate linear model categorical data methods with respect to either maximum likelihood or weighted least squares estimation and hypothesis testing procedures	

4 Specific recommendations for the examples in Section 2

In this section, some recommendations are given for the examples described in Section 2. Since their discussion is necessarily brief, these recommendations should be viewed as general guidlines concerning some of the statistical issues which underly the analysis of repeated measurements and the types of methods which can be used to deal with them. They are not intended to be fixed specifications for distinguishing between 'valid' and 'non-valid' strategies since this type of consideration is often dependent on the objectives of the researcher and the context of the application. Thus, as is the case with all aspects of statistical methodology, they should be used carefully and sensibly.

4.1 *A split-plot agricultural experiment*

For the agricultural experiment concerning fertilizers and methods of cultivation, repeated measurement (split-plot) analysis of variance and/or its non-parametric counterparts represent an appropriate strategy which may be advantageous with respect to power. However, caution should be exercised with respect to the test of treatment × condition interaction for which the multivariate approach is generally preferable. The multivariate approach may be more straightforward to implement from a practical point of view if there are missing data and/or if there are experimental design complexities such as confounding. Other discussion of this type of example is given in Anderson and Bancroft (1952), Federer (1955), Cole and Grizzle (1966), and Winer (1971).

4.2 *A growth study*

Since the baby animal growth study involves a basic observational unit, (*i.e.*, weeks) which is not randomized, some type of multivariate approach is required. Thus, the most straightforward analytical strategy with respect to implementation is the use of multivariate profile analysis of variance procedures and/or their non-parametric counterparts with mother animals as the experimental unit. However, if this is done, the data vector per experimental unit can be somewhat complicated if the research design is unbalanced with respect to the assignment of baby animals to their diets within mothers and/or if there are missing data for certain weeks. For this reason, the multivariate approach may need to be directed at within experimental unit summary measures for each baby animal diet like average growth changes for certain time intervals, average segmented regression slopes for certain time intervals, or average estimated parameter values for certain growth curve models. Since such summary measures may not necessarily have the same distributional structure from one experimental unit to another, statistical tests of significance for mother diet and mother diet × baby diet interaction should be undertaken from a randomization model point of view either with respect to the observed measurement scale or ranks. Also, statistical tests of significance for baby diet should be undertaken via non-parametric signed rank (or score) methods. Alternatively, baby diets could be compared by using the randomization structure for baby animals within mothers with respect to summary growth measures for each baby animal. Finally, if large sample sizes are under study, multivariate linear models involving weighted least squares can be used to investigate the relationship between mother diet, baby diet, and week. Other discussion of this type of example is given in Grizzle and Allen (1969), Koch (1969), Nelder and Wedderburn (1972), and Snee (1972).

4.3 *An industrial response curve study*

Since this experiment involves an observational unit (*i.e.*, time) which is not randomized, some type of multivariate approach is required. In this regard, three types of within experimental unit summary measures of interest are regression coefficients for some appropriate response curve model, principal components with respect to all time points, and a combination of the average of the observational unit data for the respective time points and modified principal components for differences among them. See Church (1966), Grizzle and Allen (1969), Morrison (1976), and Snee (1972) for further discussion of this type of example.

4.4 *A drug sensitivity study*

For the drug sensitivity study, Cochran's Q statistic represents an appropriate method of analysis since cultures have been randomly assigned to drugs on a within child basis. Alternatively, multivariate linear model methods involving maximum likelihood or weighted least squares could be used to compare drugs in terms of their estimated effects. Other dicussion of this type is given in Bhapkar and Somes (1977) and Koch *et al.* (1977).

4.5 *A multivariate study of dairy cows*

For the dairy cow example, multivariate Kruskal-Wallis methods can be used to compare treatment groups with respect to post-treatment status, pre-treatment vs post-treatment change, and post-treatment status after covariance adjustment for pre-treatment status. In addition, these methods can be applied to total numbers of diseased quarters per cow as summary measures of overall experience as well as to quarter difference scores as measures of quarter × treatment interaction. Alternatively, multivariate linear model methods involving weighted least aquares could be used to investigate the relationship of post-treatment status to treatment and quarter after adjustment for pre-treatment status. Other discussion of this type of example is given in Koch *et al.* (1978).

4.6 *A longitudinal clinical trial*

For the multi-clinic clinical trial, generalized Cochran/Mantel-Haenszel strategies can be used to combine within clinic multivariate Kruskal-Wallis or signed rank methods across clinics in a manner whose validity does not require any assumptions concerning treatment × clinic interaction (although the presence of potential interaction should be investigated in its own right because of its implications to interpretation). Otherwise, the analysis strategies for this type of situation are similar to those described for the dairy cow example in 4.5. Additional discussion of this type of example is given in Stanish *et al.* (1978).

4.7 *An agricultural product shape study*

Since this experiment involves an observational unit which is non-randomized, some type of multivariate approach is required. Two types of within experimental unit functions of interest are the average value of the diameter/shape ratios for the respective positions and modified principal components for differences among them. Finally, since normal distribution assumptions may not be appropriate for diameter/length ratio based functions, a non-parametric framework may be preferable for the evaluation of statistical significance. See Snee and Andrews (1971) for further discussion of this type of example.

4.8 *A change-over clinical trial*

For the change-over clinical trial, several different types of analysis are feasible. In particular, the data obtained during the first period can be analyzed by the same methods alluded to for the parallel clinical trial in 4.6 including non-parametric covariance adjustment for pre-treatment status. However, appropriate use of the information obtained during other periods may be somewhat complicated if the patients in the sequence groups are no longer similar at the beginning of those periods. Although this issue warrants definite concern, it does not necessarily mean that such data should be ignored (nor does it imply that multi-period research designs not be used in contexts where they have strong substantive justification even with possibly some statistical limitations), but rather that their role in analysis requires careful consideration. For this purpose, one strategy is to apply the randomization structure for the overall clinical trial to the data for the periods subsequent to the first in a manner which adjusts for patient status at their beginning via post-stratification procedures. Alternatively, if large sample sizes are under study, multivariate linear models weighted least squares methods could be used to investigate the relationship of patient health status to treatment, period, and week during period. Other aspects of this type of example are discussed in Amara *et al.* (1977), Gart (1969), Grizzle (1965), Koch (1972), and Koch *et al.* (1977).

4.9 *A comparative attitudes survey*

The political survey concerning trust of institutions involves non-randomized observational units, and thus some type of multivariate approach is required. Since such data are usually nominal or ordinal, the most appropriate analysis strategies are either multivariate non-parametric procedures for categorical data or multivariate linear model procedures. Other discussion of this type of example is given in Lehnen and Koch (1974).

4.10 *A behavioral development study*

This longitudinal study involves non-randomized observational units, and thus some type of multivariate approach is required. Since such data are usually binary, the most appropriate analysis strategies are either multivariate non-parametric procedures for categorical data or multivariate linear model procedures. Other discussion for this type of example is given in Koch *et al.* (1977) and Landis *et al.* (1979).

4.11 *A rank preference survey*

Since the tax-expenditure rank preference data involve multivariate categorical data for non-randomized observational units, their analysis should be undertaken via either multivariate non-parametric procedures or multivariate linear model procedures. However, here additional attention has to be given to the restriction that the rank of any one of the alternatives is determined from the ranks assigned to the other six. Additional aspects of this type of example are discussed in Koch *et al.* (1976).

4.12 *An observer agreement study*

For the observer agreement study, multivariate methods should be used to investigate among observer differences (*i.e.*, observer bias) since observer determinations are not randomized. Also, intraclass correlation oriented measures of observer agreement (or K-statistics) should be analyzed in such situations. Other aspects of this type of example are discussed in Landis and Koch (1977).

Acknowledgements

This research was supported in part by the U.S. Bureau of the Census (JSA 79-16) and Hoechst-Roussel Pharmaceuticals, Inc. The authors would like to thank John Gill, Oscar Kempthorne, Ronald Snee, and Robert Starbuck for comments which were helpful with respect to the preparation of this paper.

5 Reference For Topics (Some Quoted In Text)

5.1 *Repeated measurement (split-plot) analysis of variance*

Anderson, R.L. and Bancroft, T.A. (1952). *Statistical Theory in Research*. McGraw-Hill Book Company, New York.

Federer, W.T. (1955). *Experimental Design*. The Macmillan Company, New York.

Federer, W.T. (1975). The misunderstood split plot. In *Applied Statistics*, edited by R.P. Gupta. North Holland Publishing Co., Amsterdam, 9–39.

Gill, J.L. (1978). *Design and Analysis of Experiments in the Animal and Medical Sciences*. The Iowa State University Press, Ames, Iowa.

Grizzle, J.E. (1965). The two-period change-over design and its uses in clinical trials. *Biometrics*, **21**, 467–480.

Snee, R.D. and Andrews, H.P. (1971). Statistical design and analysis of shape studies. *The Journal of the Royal Statistical Society Series C 20*, 250–258.

Snee, R.D. (1972). A useful method for conducting carrot shape studies. *Journal Horticulture Science*, **47**, 267–277.

Snee, R.D. Acuff, S.K. and Gibson, J.R. (1979). A useful method for the analysis of growth studies. *Biometrics*, **35**, 835–848.
Wallenstein, S. and Fisher, A.C. (1977). The analysis of the two-period repeated measurements crossover design with applications to clinical trials. *Biometrics*, **33**, 261–269.
Winer, B.J. (1971). *Statistical Principles in Experimental Design, Second Edition*. McGraw-Hill Book Company, New York.

5.2 *Repeated measurement multivariate (profile) analysis of variance*

Anderson, T.W. (1978). Repeated measures in autoregressive processes. *Journal of the American Statistical Association*, **73**, 371–378.
Bock, R.D. (1975). *Multivariate Statistical Methods in Behavioral Research*. McGraw Hill Book Company, New York.
Bock, R.D. (1979). Univariate and multivariate analysis of variance of time-structured data. Chapter 8 in *Longitudinal Research in the Study of Behavior and Development*, edited by J.R. Nesselroade and P.B. Baltes. Academic Press, New York, 199–232.
Box, G.E.P. (1950). Problems in the analysis of growth and wear curves. *Biometrics*, **6**, 362–389.
Cole, J.W.L. and Grizzle, J.E. (1966). Applications of multivariate analysis of variance to repeated measurements experiments. *Biometrics*, **22**, 810–828.
Church, A., Jr. (1966). Analysis of data when the response is a curve. *Technometrics*, **8**, 229–246.
Danford, M.B., Hughes, H.M. and McNee, R.C. (1960). On the analysis of repeated measurements experiments. *Biometrics*, **16**, 547–565.
Davidson, M.L. (1972). Univariate versus multivariate tests in repeated measures experiments. *Psychological Bulletin*, **77**, 446–452.
Greenhouse, S.W. and Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, **24**, 95–112.
Grizzle, J.E. (1970). An example of the analysis of a series of response curves and an application of multivariate multiple comparisons. Chapter 15 in *Essays in Probability and Statistics in the Memory of S. N. Roy*, edited by R.C. Bose *et al.* University of North Carolina Press, Chapel Hill, N.C., 311–326.
Grizzle, J.E. and Allen, D.M. (1969). Analysis of growth and dose response curves. *Biometrics*, **25**, 357–381.
Grizzle, J.E., Williams, O.D. and Woolson, R.F. (1975). Multivariate analysis in biometry. *Proceedings of the 8th International Biometric Conference*, 73–84.
Guire, K.E. and Kowalski, C. (1979). Mathematical description and representation of developmental change functions in the intra- and inter-individual levels. Chapter 4 in *Longitudinal Research in the Study of Behavior and Development*, edited by J.R. Nesselroade and P.B. Baltes. Academic Press, New York, 89–110.
Huynh, H. (1978). Some approximate tests for repeated measurement designs. *Psychometrika*, **43**, 161–175.
Joreskog, K.G. (1979). Statistical estimation of structural models in longitudinal-developmental investgations. Chapter 11 in *Longitudinal Research in the Study of Behavior and Development*, edited by J.R. Nesselroade and P.B. Baltes. Academic Press, New York, 303–351.
Kornegay, E.T., Mosanghini, V. and Snee, R.D. (1970). Urea and amino acid supplementation of swine diets. *Journal of Nutrition*, **100**, 330–340.
Kowalski, C.J. and Guire, K.E. (1974). Longitudinal data analysis. *Growth*, **38**, 131–169.
Madsen, K.W. (1977). A growth curve model for studies in morphometrics. *Biometrics*, **33**, 659–669.
Morrison, D.F. (1976). *Multivariate Statistical Methods, Second Edition*. McGraw-Hill Book Company, New York.
Morrison, D.F. (1972). The analysis of a single sample of repeated measurements. *Biometrics*, **28**, 55–71.
Nelder, J.A. and Wedderburn, R.W. (1972). Generalized linear models. *Journal of the Royal Statistical Society A*, **135**, 370–384.
Poor, D D.S. (1973). Analysis of variance in repeated measures designs: Two approaches. *Psychological Bulletin*, **80**, 204–209.
Potthoff, R.F. and Roy, S.N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika*, **51**, 313–326.
Rao, C.R. (1958). Some statistical methods for comparison of growth curves. *Biometrics*, **14**, 1–17.
Rao, C.R. (1965). *Linear Statistical Inference and Its Applications*. John Wiley and Sons, Inc., New York.
Scheffe, H. (1959). *The Analysis of Variance*. John Wiley and Sons, Inc., New York.
Schwertman, N.C. (1978). A note on the Geisser-Greenhouse correction for incomplete data split-plot analysis. *Journal of the American Statistical Association*, **73**, 393–396.
Snee, R.D. (1972). On the analysis of response curve data. *Technometrics*, **14**, 47–61.
Timm, N. H. (1975). *Multivariate Analysis with Applications in Education and Psychology*. Brooks/Cole Publishing Company, Monterey, California.
Wishart, J. (1938). Growth-rate determinations in nutrition studies with the bacon pig, and their analysis. *Biometrika*, **30**, 16–28.

5.3 *Non-parametric rank methods*

Amara, I.A., Koch, G.G. and Markley, J.B. (1977). A statistical strategy for non-parametric analysis of longitudinal data from a two-period change-over design. *Statistical Computing Section Proceedings of the American Statistical Association*, 108–113.

Bhapkar, V.P. and Patterson, K.W. (1977). On some non-parametric tests for profile analyses of several multivariate samples. *Journal of Multivariate Analysis*, **7**, 265–277.

Bhapkar, V.P. and Patterson, K.W. (1978). A Monte Carlo Study of some multivariate non-parametric statistics for profile analysis of several samples. *Journal of Statistical Computation and Simulation*, **6**, 223–237.

Chatterjee, S.K. and Sen, P.K. (1970). Non-parametric tests for the multivariate multi-sample location problem. Chapter in *Essays in Probability and Statistics in memory of S. N. Roy*, edited by R.C. Bose *et al.* University of North Carolina Press, Chapel Hill, N.C., 197–229.

Conover, W.J. (1971). *Practical Nonparametric Statistics*. John Wiley and Sons, Inc., New York.

Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*, **32**, 675–699.

Gerig, T. (1969). A multivariate extension of Friedman's test. *Journal of the American Statistical Association*, **64**, 1595–1608.

Ghosh, M., Grizzle, J.E. and Sen, P.K. (1973). Nonparametric methods in longitudinal studies. *Journal of the American Statistical Association*, **68**, 29–36.

Gillings, D.B., Symons, M.J. and Donelan, M.M. (1978). Strategies for the analysis of repeated measurements designs involving ordinal categorical data. Institute of Statistics Mimeo Series No. 1215. Department of Biostatistics, University of North Carolina, Chapel Hill.

Hollander, M. and Wolfe, D.A. (1973). *Nonparametric Statistical Methods*. John Wiley and Sons, Inc., New York.

Koch, G.G. (1969). Some aspects of the statistical analysis of 'split plot' experiments in completely randomized layouts. *Journal of the American Statistical Association*, **64**, 485–505.

Koch, G.G. (1970). The use of non-parametric methods in the statistical analysis of a complex split plot experiment. *Biometrics*, **26**, 105–128.

Koch, G.G. (1972). The use of non-parametric methods in the statistical analysis of the two-period change-over design. *Biometrics*, **28**, 577–584.

Koch, G.G. and Sen, P.K. (1968). Some aspects of the statistical analysis of the mixed model. *Biometrics*, **24**, 27–47.

Kruskal, W. H. and Wallis, W. A. (1953). Use of ranks in one criterion variance analysis. *Journal of the American Statistical Association*, **46**, 583–621.

Puri, M.L. and Sen, P.K. (1971). *Nonparametric Methods in Multivariate Analysis*. John Wiley and Sons, Inc., New York.

Quade, Dana. (1967). Rank analysis of covariance. *Journal of the American Statistical Association*, **62**, 1187–1200.

Sen, P.K. (1968). On a class of aligned rank order tests in two-way layouts. *Annals of Mathematical Statistics*, **39**, 1115–1124.

Singer, B. (1979). *Distribution-free methods for non-parametric problems: A classified and selected bibliography*. The British Psychological Society, Leicester, England.

Zerbe, G.O. and Walker, S.H. (1977). A randomization test for comparison of groups of growth curves with different polynomial design matrices. *Biometrics*, **33**, 653–657.

5.4 *Categorical data methods – multivariate non-parametric procedures*

Bhapkar, V.P. and Somes, G.W. (1977). Distribution of Q when testing equality of matched proportions. *Journal of the American Statistical Association*, **72**, 658–661.

Cochran, W. (1950). The comparison of percentages in matched samples. *Biometrika*, **37**, 256–266.

Fleiss, J.L. (1973). *Statistical Methods for Rates and Proportions*. John Wiley and Sons, Inc., New York.

Gart, J.J. (1969). An exact test for comparing matched proportions in crossover designs. *Biometrika*, **56**, 75–80.

Hettmansperger, T. P. and McKean, J.W. (1973). On testing for significant change in c × c tables. *Communications in Statistics*, **2**, 551–560.

Koch, G.G. (1977). The interface between statistical methodology and statistical practice. *Social Statistics Section, 1977 Proceedings of the American Statistical Association*, 205–214.

Landis, J.R., Heyman, E.R. and Koch, G.G. (1978). Average partial association in three-way contingency tables: A review and discussion of alternative tests. *International Statistical Review*, **46**, 237–254.

Mantel, N. (1963). Chi-square tests with one degree of freedom; extensions of the Mantel-Haenszel procedure. *Journal of the American Statistical Association*, **58**, 690–700.

Mantel, N. and Byar, D.P. (1978). Marginal homogeneity, symmetry, and independence. *Communications in Statistics* A7(10) 953–976.

Nam, J. (1971). On two tests for comparing matched proportions. *Biometrics*, **27**, 945–959.

Somes, G. W. and Bhapkar, V.P. (1977). A simulation study on a Wald statistic and Cochran's Q statistic for stratified samples. *Biometrics*, **33**, 643–651.

5.5 *Categorical data methods – multivariate linear model procedures*

Bishop, Y.M.M., Fienberg, S.E. and Holland, P.W. (1975). *Discrete Multivariate Analysis*. The MIT Press, Cambridge.

Gokhale, D.V. and Kullback, S. (1978). *The Information in Contingency Tables*. Marcel Dekker, Inc., New York.

Imrey, P.B., Johnson, W.D. and Koch, G.G. (1975). An incomplete contingency table approach to paired-comparison experiments. *Journal of the American Statistical Association*, **71**, 614–623.

Koch, G.G. and Bhapkar, V.P. (1980). Chi-square tests. To appear in *Encyclopedia of Statistical Sciences*, edited by Norman L. Johnson and Samuel Kotz. John Wiley and Sons, Inc., New York.

Koch, G.G., Freeman, J.L. and Lehnen, R.G. (1976). A general methodology for the analysis of ranked policy preference data. *International Statistical Review*, **44**, 1–28.

Koch, G.G., Gillings, D.B. and Stokes, M.E. (1980). Biostatistical implications of design, sampling and measurement to the analysis of health science data. *Annual Review of Public Health*, **1**, 163–225.

Koch, G.G., Grizzle, J.E., Semenya, K. and Sen, P.K. (1978) Statistical methods for evaluation of mastitis treatment data. *Journal of Dairy Science*, **61**, 829–847.

Koch, G.G., Imrey, P.B. and Reinfurt, D.W. (1972). Linear model analysis of categorical data with incomplete response vectors. *Biometrics*, **28**, 663–692.

Koch, G.G., Landis, J.R., Freeman, J.L., Freeman, D.H. and Lehnen, R.G. (1977). A general methodology for the analysis of experiments with repeated measurement of categorical data. *Biometrics*, **33**, 133–158.

Koch, G.G. and Reinfurt, D.W. (1971). The analysis of categorical data from mixed models. *Biometrics*, **27**, 157–173.

Landis, J.R. and Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, **33**, 159–175.

Landis, J.R. and Koch, G.G. (1979) The Analysis of Categorical Data in Longitudinal Studies of Behavioral Development. Chapter 9 in *Longitudinal Research in the Study of Behavior and Development*, edited by J.R. Nesselroade and P.B. Baltes. Academic Press, New York, 233–261.

Layard, M.W.J. and Arvesen, J.N. (1978). Analysis of Poisson data in crossover experimental designs. *Biometrics*, **34**, 421–428.

Lehnen, R.G. and Koch, G.G. (1974). Analyzing panel data with uncontrolled attrition. *The Public Opinion Quarterly*, **38**, 40–56.

Lehnen, R.G. and Koch, G.G. (1974). The analysis of categorical data from repeated measurement research designs. *Political Methodology*, **1**, 103–123.

Plackett, R.L. (1974). *The Analysis of Categorical Data*. London: Griffin.

Stanish, W.M., Gillings, D.B. and Koch, G.G. (1978). An application of multivariate ratio methods for the analysis of a longitudinal clinical trial with missing data. *Biometrics*, **34**, 305–317.

Résumé

Un caractère commun à bien des recherches statistiques est la collecte de données faite sur des groupes d'unités expérimentales dont chacune est observée sous deux conditions ou davantage. De telles études sont appelées généralementsoit des expériences 'split-plot', soit des expériences à mesures répétées. Le présent article a pour objet de passer en revue les stratégies générales de la statistique pour analyser des données provenant de ces types de plans de recherches. A cet effet, l'attention se porte en premier dans deux directions fondamentales. L'une est la nature des processus de 'randomization' affectant les données obtenues parmi les unités expérimentales ou bien à l'intérieur de celles-ci. L'autre direction, c'est à quel niveau se situe l'échelle de mesure qu'on peut qualifier de nominal, d'ordinal ou de scalaire. Ce cadre de travail sert alors de base pour discuter des méthodes statistiques en compétition, telles que l'analyse de variance des mensurations répétées, leur analyse de variance multivariée, et leurs contreparties non-paramétriques pour des données ayant la forme de rangs ou bien de catégories, – discussion faite à la fois dans le cas général et sur des exemples de classes déterminées. Pour finir, on donne une bibliographie de références choisies, concernant des méthodes et celles qui en sont proches.