

Permutation procedures for ANOVA, Regression and PCA

by

Christine Storm

Submitted in partial fulfilment of the requirements for the degree

Master of Science (Mathematical Statistics)

In the Faculty of Natural & Agricultural Sciences

University of Pretoria

Pretoria

2012

Declaration

I, the undersigned, hereby declare that this dissertation, which I hereby submit for the degree Master of Science at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution.

Signature _____

Date _____

Acknowledgements

I would like to sincerely thank my supervisor Dr. L Fletcher for her extraordinary efforts to my benefit throughout my tenure as a MSc student in the Department of statistics at the University of Pretoria. Her knowledge and guidance during my research was indispensable.

Summary

Parametric methods are effective and appropriate when data sets are obtained by well-defined random sampling procedures, the population distribution for responses is well-defined, the null sampling distributions of suitable test statistics do not depend on any unknown entity and well-defined likelihood models are provided for by nuisance parameters.

Permutation testing methods, on the other hand, are appropriate and unavoidable when distribution models for responses are not well specified, nonparametric or depend on too many nuisance parameters; when ancillary statistics in well-specified distributional models have a strong influence on inferential results or are confounded with other nuisance entities; when the sample sizes are less than the number of parameters and when data sets are obtained by ill-specified selection-bias procedures. In addition, permutation tests are useful not only when parametric tests are not possible, but also when more importance needs to be given to the observed data set, than to the population model, as is typical for example in biostatistics.

The different types of permutation methods for analysis of variance, multiple linear regression and principal component analysis are explored. More specifically, one-way, two-way and three-way ANOVA permutation strategies will be discussed. Approximate and exact permutation tests for the significance of one or more regression coefficients in a multiple linear regression model will be explained next, and lastly, the use of permutation tests used as a means to validate and confirm the results obtained from the exploratory PCA will be described.

Contents

| | | |
|----|--|----|
| 1. | <i>Introduction</i> | 3 |
| 2. | <i>Notation and Abbreviations</i> | 5 |
| 3. | <i>Conditionality and Exchangeability</i> | 6 |
| 3 | <i>Randomization and Permutation</i> | 9 |
| 4 | <i>When Permutation is Appropriate</i> | 11 |
| 5 | <i>The Beginnings of Permutations</i> | 13 |
| | 5.1.1 1920–1939 | 13 |
| | 5.1.2 1940–1959 | 16 |
| 6 | <i>Computational Aspects</i> | 17 |
| | 6.1.1 1960–1979 | 17 |
| | 6.1.2 1980–1999 | 18 |
| | 6.1.3 2000–2012 | 20 |
| 7 | <i>Optimal Procedures</i> | 21 |
| 8 | <i>Analysis of Variance</i> | 26 |
| | 8.1 Introduction | 26 |
| | 8.2 One-Way Analysis of Variance..... | 27 |
| | 8.2.1 <i>The Parametric Approach</i> | 27 |
| | 8.2.2 <i>The Permutation Approach</i> | 29 |
| | 8.3 Two-Way Analysis of Variance..... | 31 |
| | 8.3.1 <i>The Parametric Approach</i> | 31 |
| | 8.3.2 <i>Permutation of Raw Data</i> | 33 |
| | 8.3.3 <i>Permutation under the Reduced Model</i> | 35 |
| | 8.3.4 <i>Permutation of the Full Model</i> | 36 |
| | 8.3.5 <i>Permutation for an Exact Test</i> | 37 |
| | 8.4 Three-Way Analysis of Variance | 42 |
| | 8.5 Simulation Study | 43 |
| | 8.6 Numerical Example | 45 |
| 9 | <i>Multiple Linear Regression</i> | 47 |
| | 9.1 Introduction | 47 |
| | 9.2 The Parametric Approach..... | 48 |
| | 9.3 Permutation of Raw Data | 50 |
| | 9.3.1 <i>The Manly Method</i> | 50 |

| | | |
|-------|--|-----|
| 9.4 | Permutation under the Reduced Model | 52 |
| 9.4.1 | <i>The Freedman and Lane Method</i> | 52 |
| 9.4.2 | <i>The Kennedy Method</i> | 53 |
| 9.5 | Permutation under the Full Model | 56 |
| 9.5.1 | <i>The Ter Braak Method</i> | 56 |
| 9.5.2 | <i>The Tantawanich Method</i> | 58 |
| 9.6 | Permutation for an Exact Test | 60 |
| 9.6.1 | <i>The Kherad-Pajouh and Renaud Method</i> | 60 |
| 9.7 | Simulations | 66 |
| 10 | <i>Principal Component Analysis</i> | 72 |
| 10.1 | Introduction | 72 |
| 10.2 | The Principal Component Procedure | 73 |
| 10.3 | The Permutation Approach | 74 |
| 10.4 | Two Permutation Strategies | 75 |
| 10.5 | Numerical Example | 76 |
| 11 | <i>Appendix</i> | 80 |
| 12 | <i>References</i> | 84 |
| 13 | <i>Appendix</i> | 89 |
| 13.1 | Code for Section 8.3 | 89 |
| 13.2 | Code for Section 8.5 | 95 |
| 13.3 | Code for Section 9.7 | 99 |
| 13.4 | Code for Section 10.5 | 104 |

1. Introduction

Permutation tests are currently the gold standard against which conventional parametric tests are tested and evaluated. In this document, permutation statistical methods are introduced, a historical chronology of the development of permutation methods is provided and the advantages of permutation methods are detailed. The different types of permutation methods are also described for analysis of variance, multiple linear regression and principal component analysis. These permutation methods are then compared to the traditional parametric tests using examples and simulations.

The population model assumes random sampling from one or more specified population. Under the population model, the level of statistical significance that results from applying a statistical test to the results of an experiment or a survey corresponds to the frequency with which the null hypothesis would be rejected in repeated random samplings from the same specified population. Because repeated sampling of the true population is usually impractical, it is assumed that the sampling distribution of the test statistics under repeated random sampling conforms to an assumed theoretical distribution, such as the normal distribution. The size of the test, for example 0.05, is the probability under a specified null hypothesis that repeated outcomes based on random samples of the same size are equal to or more extreme than the observed outcome. In the population model, assignment of treatment to subjects is viewed as fixed with a stochastic element taking the form of the error that would vary if the experiment was repeated. Probabilities are then calculated based on the potential outcomes of conceptual repeated draws of these errors.

With the permutation approach, a test statistic is computed for the observed data, then the data are permuted over all possible arrangements of the observed data and the test statistic is computed for each likely arrangement. An ordered sequence of n exchangeable objects yields $n!$ equally likely arrangements of the n objects. The proportion of arrangements with test statistic values equal to or more extreme than the observed case yields the probability of the observed test statistic. Probabilities are then calculated according to all outcomes associated with assignments of treatments to subjects for each case.

Permutation tests differ from traditional parametric tests in several ways. Permutation tests are data dependent, in that all the information required for analysis is contained within the observed data set. Permutation tests do not assume any underlying theoretical distribution. Permutation tests do not depend on the assumptions associated with traditional parametric tests, such as normality and homogeneity. Permutation tests provide probability values based on the discrete permutation distribution of equally likely test statistic values, rather than an approximate probability value based on a theoretical distribution, such as a normal.

2. Notation and Abbreviations

ANOVA: Analysis of variance

PCA : Principal Component Analysis

i.i.d : independent and identically distributed

$N(\mu, \sigma^2)$: Gaussian or normal variable with mean μ and variance σ^2

OLS : ordinary least squares

\sim : distributed as

n : the (finite) sample size

$tr()$: the trace of a matrix

X : a univariate random variable

\mathbf{X} : a multivariate variable or a sample of n units, $\mathbf{X} = \{X_i, i = 1, \dots, n\}$

\mathbf{X}^* : a permutation of \mathbf{X}

F^{M*} : permutation statistic for Manly (1991, 1997, 2007)

F^{E*} : permutation statistic for Edgington (2007)

F^{SW*} : permutation statistic for Still and White (1981)

F^{J*} : permutation statistic for Jung et al. (2006)

F^{FL*} : permutation statistic for Freedman and Lane (1983)

F^{K*} : permutation statistic for Kennedy (1995)

F^{TB*} : permutation statistic for Ter Braak (1992)

F^{T*} : permutation statistic for Tantawanich (2006)

F^{KR*} : permutation statistic for Kherad-Pajouh and Renaud (2010)

3. *Conditionality and Exchangeability*

For most problems of hypothesis testing, the observed data set $P \in \mathcal{Y} = \{y_1, \dots, y_n\}$ is usually obtained by an experiment performed n times on a population variable X . For the purposes of analysis, the data set x is generally partitioned into groups or samples, according to the treatment levels of the experiment. For any general testing problem, under the null hypothesis, which assumes that data comes from only one (with respect to groups) unknown population distribution \mathcal{P} , the whole set of observed data x is considered to be a random sample.

Pesarin (2001) defines nonparametric distributions as follows: A family of distributions \mathcal{P} is said to behave non-parametrically when we are not able to find a parameter θ , belonging to a known finite-dimensional parameter space Θ and \mathcal{P} , in a sense that each member of \mathcal{P} cannot be identified by only one member of Θ and vice versa.

This definition by Pesarin (2001) includes families of distributions which are either unspecified or specified, except for an infinite number of unknown parameters. All nonparametric families \mathcal{P} which are of interest in permutation analysis are assumed to be sufficient in such a way that if x and $x\emptyset$ are any two points, then $x \neq x'$ implies $f_p(x) \neq f_p(x')$ for at least one $P \in \mathcal{P}$, except for points with null density. The characterisation of a family \mathcal{P} as being nonparametric essentially depends on the knowledge we assume about it. When we assume that the underlying family \mathcal{P} contains all continuous distributions, then the data set x is sufficient. By sufficiency, it means that x and f_p are said to contain essentially the same amount of information with respect to P . They are equivalent for inferential purposes.

The same conclusion is obtained if the sample distribution is assumed to be invariant with respect to permutations of the arguments of x . This happens when the assumption of independence for observable data is replaced by that of exchangeability: $f(x_1, \dots, x_n) = f(x_{u_1^*}, \dots, x_{u_n^*})$, where (u_1^*, \dots, u_n^*) is any permutation of $(1, \dots, n)$. The data sets under the null hypothesis is always contain a finite number of points, as n is finite (Good; 2005).

Permutation tests are conditional statistical procedures. Under the null hypothesis and assuming exchangeability, the conditional probability distribution for whatever underlying population distribution $P \in \mathcal{P}$, is

$$P(\mathbf{x}^* = \mathbf{x}') = \frac{\text{No. of the same or more extreme outcomes as that observed}}{\text{Total no. of possible outcomes}}$$

which is P -independent. If there are no ties in the data set then the conditional probability becomes $1/n!$. Thus $P(\mathbf{x}^* = \mathbf{x}')$ is uniform on the permutation sample space for all $P \in \mathcal{P}$. In the case of the classical t -, F - and χ^2 -tests, it is not a point probability but consists of the probability contained in the tails of the frequency distribution. How the null hypothesis is formulated depends on the outcome defined by the investigator, the design of the experiment and the scale of measurement used to obtain the experimental values. For instance, in the case of two sets of observations measured on an interval scale, the outcome could be a difference between arithmetic means, geometric means, medians, mid-ranges, mean-ranks, proportions or even variances. If there are more than two sets of observations, a conventional test would be classical ANOVA. Thus the equivalent of the F -statistic can be permuted. When measurements have been made on a nominal scale, test statistics such as Pearson's χ^2 , the likelihood-ratio or the odds-ratio can be permuted (Good; 1994).

The conditioning of permutation procedures allows permutation inferences to be invariant with respect to P in the null hypothesis. Some authors prefer to give them the name of invariant tests. However, according to Ludbrook (1994), because of this invariance property, permutation tests are distribution-free and nonparametric.

The condition of exchangeability on sufficient statistics provides permutation tests with good general properties. One of these is that, when exchangeability is satisfied in the null hypothesis, permutation tests are always exact procedures (Berry and Mielke; 1985). If data come from continuous distributions, so that the probability of finding ties in the data set is zero, the rejection probability under the null hypothesis is invariant with respect to the observed data set \mathbf{x} . Thus rejection regions are similar to the conventional parametric region.

When data come from non-continuous distributions, Berry and Mielke (1985) show that the similarity property is only asymptotically valid.

Permutation inferences are proper with observational data, which sometimes are called non-experimental, and with well-designed sampling procedures. However, well-designed sampling procedures are quite rare even in most experimental problems (Anderson; 2003). For instance, if we want to investigate the effects of a drug on rats, the units to be treated are usually not randomly chosen from the population of all rats, but are selected in some way among those available in a laboratory and are randomly assigned to the established treatment levels. The same occurs in most clinical trials, in which some of the patients present in a hospital are randomly assigned to one of the pre-established treatment levels. In one sense, the concept of random sampling is rarely achieved in real applications because, for various reasons, real samples are commonly obtained by selection-bias procedures (Good; 2005). This implies that parametric tests, being based on the concept of random sampling, are rarely applicable in real situations. Additionally, because of the similarity and unbiasedness properties, permutation solutions allow for relaxation of most of the common assumptions needed by parametric counterpoints, such as the existence of mean values and variances, and the homoscedasticity of responses in the alternative hypothesis. This is why permutation inferences are important for both theoretical and the application aspects (Gonzales and Manly; 1998).

Within the assumption of exchangeability in the null hypothesis, permutation conditional inferences always have a clear interpretation, whereas extensions to the underlying parent population should be carried out and interpreted carefully. These extensions and associated interpretations are generally easy and correct when data are designed from random sampling techniques from a given population. Of course, if they are collected by selection-bias procedures, these extensions may sometimes be ambiguous and misleading. Many authors such as Good (2005) and Edgington (1995) have emphasized these aspects. One of these relates to the fact that reference null distributions of ordinary parametric tests are explicitly based on the concept of infinitely repeated and well-designed random sampling from a given well-specified population. This existence of this random sampling is often merely virtual. As it occurs in many experimental problems, it is often to unrealistic to assume that treatment

does not influence scale coefficients or other aspects of interest, so that standard parametric solutions may become improper.

Conversely, when exchangeability may be assumed under the null hypothesis the reference null distributions of permutation tests always exist, because at least in principle, they are obtained by considering all permutations of available data. In addition, permutation comparisons of means do not require homoscedasticity. For these reasons, permutation inferences generally have a natural interpretation and ordinary parametric tests are considered to be rarely applicable to real problems (Anderson; 2003).

3 Randomization and Permutation

The rationale behind randomization tests is to consider a problem in which a test statistic T is being used to test a null hypothesis. Suppose that a suitable shuffling of the data produces a new configuration of the data. From the new configuration the statistic T^* can be calculated. Under the null hypothesis, this configuration can be viewed as equally likely. By repeating this for all possible shuffles one can assess the extent to which T^* is unusual and thereby accept or reject the null hypothesis.

Those occupied in nonparametric statistics, such as Dallal (1988), recognise that this principle lies behind rank tests, sometimes referred to as rank permutation tests. There are a finite number of possible outcomes in a rank-test statistic, calculated by permuting the ranks of relevant variables observations. If the value actually obtained is unusual relative to these possible values, the null hypothesis is cast in doubt.

In 1935 Fisher described a way of comparing the means of randomized pairs of observations by permutation and was able to perform it exactly on a set of Charles Darwin's data on plant growth, even though 32,768 permutations were involved. However, it is extremely laborious to do these tests by hand, so in the pre-computer era only a few statisticians took them seriously (Kempthorne; 1955). Now that permutation tests are easily achievable with the

advancement in computer speeds, there has been renewed interest in them and in the randomization model of inference (Manly; 1991). Statisticians such as Fisher and Kempthorne knew, from first-hand experience, that in agricultural research random samples are not drawn from populations. In their experiments, plant varieties or different fertilizers were assigned to blocks of land within a particular field by a process of randomization. This field was not a random sample of the global population of fields, or even of fields of any definable category.

In most experimental cases units are randomly assigned to symbolic treatment levels (groups or samples) so that, under the null hypothesis, observed data appear to have been randomly assigned to these levels. Based on this notion of randomization, authors such as Pitman (1937), Kempthorne (1955) and Kennedy (1995) prefer to use the term randomization tests.

The term permutation test is thus preferable to authors such as Pesarin (2001); Good (2005) and Mielke et al. (2001) because it is closer to the true state of things. A sufficient condition for properly applying permutation tests is that the null hypothesis implies that observed data are exchangeable with respect to groups. For instance, in a symbolic experiment where a variable is observed in male and female groups of a given kind of animal, the notion of randomization is difficult to apply exactly. This is because there is no way that gender can be randomly assigned to units. Instead, the permutation idea is much more natural because, under the null hypothesis of no distributional difference due to gender, we are led to assume that observed data may be indifferently assigned to either males or females. The greater emphasis on the notion of randomization tests is because, under the null hypothesis, it is generally easier and more natural to justify the assumption of exchangeability for experimental data than for observed data.

When the exchangeability property is not satisfied or cannot be assumed under the null hypothesis, then both parametric and permutation inferences are generally not exact. In these cases, especially when even approximate solutions are difficult to obtain, it may be useful to employ bootstrap techniques, which are less demanding in terms of assumptions and may be effective at least for exploratory purposes.

The nonparametric bootstrap and permutation tests are two different approaches to the same type of problem. Whereas permutation tests are used to establish the distribution of a particular statistic under a specific null hypothesis, the bootstrap establishes the sampling distribution of a statistic based on an observed data set (Good; 2000). The bootstrap simulates resampling from the population by randomly drawing, with replacement, new samples from the observed sample. The values of the statistic of interest in the bootstrap samples form the sampling distribution from which confidence intervals can be computed. If a bootstrap confidence interval does not contain the value assumed under the null hypothesis, the observed statistic is concluded to be statistically significant. Both permutation and bootstrap methods require that the observations be independent. Both methods also require that the observations be drawn from populations in which, under the null hypothesis, a specific parameter is the same across all the populations. However, an additional assumption for permutation procedures is that observations are exchangeable, or their joint distribution be the same. The advantage of permutation tests over bootstrap is that they render exact p-values. Bootstrap procedures only provide exact p-values for very large samples. Another advantage is that the simulation of a null distribution is more closely related to traditional hypothesis testing (Edgington; 2007).

4 When Permutation is Appropriate

Parametric methods are effective and appropriate when data sets are obtained by well-defined random sampling procedures, the population distribution for responses is well-defined, the null sampling distributions of suitable test statistics do not depend on any unknown entity and well-defined likelihood models are provided for the nuisance parameters. (Gozalez and Manly; 1998).

Conversely, according to Pesarin (2001), permutation testing methods are appropriate and unavoidable when:

- Distribution models for responses are nonparametric.
- Distribution models are not well specified.

- Distribution models depend on too many nuisance parameters.
- Ancillary statistics in well-specified distributional models have a strong influence on inferential results.
- Ancillary statistics in well-specified models are confounded with other nuisance entities.
- Asymptotic null sampling distributions depend on unknown entities.
- Sample sizes are less than the number of response variables.
- Sample data comes from finite populations or sample sizes are smaller than the number of parameters.
- In multivariate problems, some variables are categorical and others quantitative.
- Multivariate alternatives are subject to order restrictions.
- In multivariate problems, component variables have different degrees of importance.
- Data sets are obtained by ill-specified selection-bias procedures.

In addition, permutation tests are useful not only when parametric tests are not possible, but also when more importance needs to be given to the observed data set, than to the population model. For example, when assessing the reliability of cars, the owner may be mostly interested in his own car or fleet of cars if he has more than one, as he is responsible for all reliability maintenance costs related to his car(s), thus giving rise to a permutation assessment. From another view point, the car manufacturer, whose reputation and warranty costs are related to the whole set of similar cars already produced, may be mostly centred on a sort of average behaviour, giving rise to a parametric assessment related to the whole car population.

Thus, both permutation and parametric points of view are important and useful in real problems, because there are situations, such as that of the car owner, in which we may be interested in permutation tests, whereas there are others, such as the car manufacturer, in which we may be interested in parametric inferences.

However, Gozalez and Manly (1998) have shown that permutation methods play a central role as they allow for extensions of parametric testing. For example, when the car manufacturer wants to compare the means of two of more populations from various car

brands, an analysis of variance can be performed. If the data does not satisfy the assumptions of parametric tests, then the analysis of variance is not valid. Thus a permutation test can be used to confirm these results.

5 The Beginnings of Permutations

5.1.1 1920–1939

The earliest indications of permutations tests appeared in a 1923 article by Neyman. In this article, Neyman introduced a model for the analysis of field experiments for the purpose of comparing a number of crop varieties. This Polish article was unknown by people working on permutation tests until 1990, when it was translated. In 1927, Geary first used an exact analysis to demonstrate the utility of asymptotic approaches for data analysis in an investigation of the properties of correlation and regression in finite populations.

Like Geary (1927), Eden and Yates (1933) used permutation methods to compare a theoretical distribution to an empirical distribution. Eden and Yates (1933) examined height measurements of wheat grown in eight blocks, each consisting of four sub-blocks of eight plots. The observations were collapsed into four treatments randomly applied to four sub-blocks in each block. The experimental data consisted of four treatment groups and four treatment blocks for a total of $(4!)^7 = 4.59E+09$ possible arrangements. Eden and Yates (1933) chose a sample of 1,000 of these arrangements at random and generated a table showing the simulated probabilities generated by the random sample and the theoretical equivalent to the probability values based on the normality assumption. The simulated and theoretical probabilities were compared by a χ^2 – goodness-of-fit test and were found to be in close agreement, supporting the assumption of normality.

In 1934, Fisher presented a paper describing the logic of a permutation test to the Royal Statistical Society. Fisher did not expressly discuss permutation tests, but instead used the binomial distribution to arrive at an exact probability for a 2×2 contingency table. The

purpose of this example was to illustrate that for small samples, exact tests are possible; thereby eliminating the need for estimation. This is indicative of an early understanding of the superiority of exact probability values computed on known discrete distributions over approximations based on theoretical distributions.

In Fisher's 1935 paper, *The Design of Experiments*, Fisher expressed the usefulness of the permutation approach to obtain exact probabilities. This famous text set the concept of permutation tests into motion. In what Fisher termed a hypothetical experiment in *The Design of Experiments*, Fisher described a woman who claimed to be able to tell the difference between tea with milk added first and tea with milk added afterwards. He concocted an experiment whereby the woman sampled eight cups of tea, four of each type, and identified the point at which the milk had been added – before the tea or after. Fisher then outlined the chances of the woman being correct merely by guessing, based on the number of trials; in this case eight cups of tea.

Fisher provided a second hypothetical discussion of permutation tests in the 1935 *Design of Experiments*, describing a way to compare the means of randomized pairs of observations by permutation. In this case Fisher carried the example through, calculating test statistics for all possible pairs of the data. For this example, Fisher considered data from Charles Darwin on fifteen pairs of planters containing *Zea Mays* seeds in similar soils and location. The heights were to be measured when the plants reached a given age. Fisher calculated the exact probability values for the $2^{15} = 3.28E+04$ possible arrangements of the data, based on the null hypothesis of no difference between self-fertilized and cross-fertilized plants. The exact probability value was calculated as the proportion of values whose differences were as extreme, or more extreme than, the observed value.

Fisher's 1936 article *The coefficient of racial likeness* provided an alternative explanation of how permutation tests work. Without calling the technique a permutation test, Fisher described a shuffling procedure for analysing data. His description began with two groups of $n = 100$ members each and a measurement of interest on each member of the two groups. The measurements were recorded on 200 cards, shuffled, and then divided at random into two groups of 100 each, a division that could be repeated in an enormous, but finite and conceptually calculable number of ways. A consideration of all possible arrangements of the

pairs of cards would provide a solution for determining if the random samples could have been drawn from the same population.

In 1936, Hotelling and Pabst calculated exact probabilities for small samples of ranked data using permutation methods. The article utilized the calculation of a probability that incorporated all permutations of the data. The assumptions were that under the null hypothesis, all permutations were equally likely. The probability for any particular value was calculated as a proportion of the number of permutations equal to, or more extreme than, the value obtained from the observed data. While earlier works demonstrated permutations tests, the article of Hotelling and Pabst introduced more extensive work on small data sets as well as the introduction of the notation $n!$. Thus, this 1936 article may be the first example that specifically detailed the method of calculating a permutation test using all possible arrangements of the data (Berry et al.; 2011).

Fisher, however, continued to be influential in the discussion of permutation methods. Welch described Fisher's inference to an exact probability and noted that although the calculations would be lengthy, the result would be a hypothesis test that was free of assumptions about the data. Pearson also referenced the Fisher text in his consideration of randomizations with the lady tasting tea, but as with Fisher, neither Welch nor Pearson fully explained the technique. It was not until 1937 that a series of articles by Pitman explicitly discussed the permutation approach for statistical analysis. These articles extended permutation methods to include data that were not amenable to ranking.

In his 1937 paper, Pitman stated that the objective of the paper was to devise valid tests of significance which involves no assumptions about the forms of the population sampled, and second noted that the idea underlying permutation tests seem to be explicit in all of Fisher's writings. Pitman further developed the permutation approach for the correlation coefficient which makes no assumptions about the population sampled, and then proposed a permutation test for the analysis of variance which involves no assumptions of normality.

In a 1938 article, on *Tests for homogeneity*, Welch advocated calculating exact values on a limited population before moving into an examination of the moments of an infinite population. Welch continued with an example of an exact calculation and further concluded

that if the variances of different samples are markedly different, normal theory could badly underestimate significant differences that might exist. However, an exact test being free from the assumptions usually associated with asymptotic statistical tests, had no such limitation.

McCarthy also argued for the use of a permutation test as a first approximation, before considering the data via an asymptotic distribution in 1939. Kendall incorporated exact probabilities utilizing the entire universe of permutations in the construction of τ , a new measure of rank correlation. In 1939, Kendall et al. utilized permutations in their discussion of Spearman's rank order correlation coefficient and exact probabilities up to $n=10$ for Spearman's rank order correlation coefficient. The probabilities were based on their relative frequencies in the $n!$ permutations of one ranking against another. This brought about the publication of tables in the 1940s for statistics with small sample sizes. These tables that employed permutations for the calculations of exact probabilities were primarily for rank tests.

5.1.2 1940–1959

The period between 1940 and 1950 brought about a rise in the work of nonparametric rank tests, publications include the Kendall rank order correlation, the Friedman two-way analysis of variance by ranks, the Wilcoxon rank sum test developed simultaneously by Wilcoxon, Mann and Whitney and the Kruskal-Wallis one-way analysis of variance by ranks. Permutation methods were employed to generate tables of exact probabilities for small samples. Theoretical work on permutation tests did continue in the 1940s and 1950s, but a theme that was commonly repeated was the conversion of data into ranks to simplify tedious computations. In 1952, Hoeffding investigated the power of a family of nonparametric tests based on permutations of observations, finding the permutation tests to be asymptotically as powerful as the related parametric tests. In 1955 Kempthorne described the use of randomization in experimental designs and how randomization permits evaluation of the experimental. Included were the completely randomized design, randomized blocks, and Latin squares. Two years later, in 1957, Dwass continued the general theme of computational difficulties for permutation tests, even with small samples. Dwass further recommended taking a random sample of all possible permutations for a two-sample test and making the

decision to accept or reject the null hypothesis on the basis of these random permutations only.

6 Computational Aspects

6.1.1 1960–1979

One reason why permutation tests are not better known is that as the size and number of randomized groups increase, there is a steep increase in the number of possible permutations of the data. For instance, when there are two independent groups, the number of possible permutations is given by the expression $(n_1 + n_2)! / n_1! n_2!$. When two groups of measurements are made on the same group, the number of possible permutations is 2^n . Thus, even when the small groups are analysed, the number of possible permutations can run to many millions. For example, two independent groups each of size 10 has $(10+10)! / 10! 10! = 1.85\text{E}+05$ possible permutations. When these two measurements are made on 20 experimental units one has $1.05\text{E}+06$ possible permutations.

If investigators are to perform permutation tests, they must have access to computers that have the capacity to perform them within a reasonable space of time. It thus has taken the development of high-speed computers for permutation tests to achieve their potential. (Berry et al.; 2011). The parallel development of permutation tests and computers is an essential part of the chronology of permutation methods.

In the period prior to 1960, computers were large, slow, and expensive, and in large part their use was limited to military and industrial applications. Computers of this era could fill up an entire warehouse. In the 1960s, mainframe computers became widely available to researchers at major research universities. By the end of the 1970s, personal computers, although not common, were available to many researchers. The speed of computers increased greatly in the 1960s and 1970s and this paved the way for the development of permutation tests (Good; 2005).

Permutation tests depend on efficient generation of permutation sequences. In the 1960s and 1970s, many algorithms were presented for the generation of permutation sequences, each aiming at increased speed and efficiency. In 1979, Mielke et al. introduced multi-response permutation procedures (MRPP), the first statistics designed especially for permutation methods, in contrast to permutation alternatives to conventional tests. The use of Euclidean distances, rather than squared Euclidean distances provided exceedingly robust, distribution-free, Euclidean-based permutation alternatives to experimental designs that normally employed conventional ANOVA or MANOVA analysis.

Researchers were focused on defining efficient methods for computing probability values in the 1960s and 1970s. Existing inefficiencies were largely due to inadequate numerical algorithms, low computer clock speeds, small core memories and inefficient data transfer. Mielke et al. (1979) introduced moment approximation permutation procedures whereby implementation of symmetric functions based on finite populations, provided the exact first three moments of a continuous distribution that approximated the discrete permutation distribution. The moment approximation permutation procedure immediately eliminated many of the computing difficulties that had inundated the computation of permutation values, provided an approximation to the underlying permutation distribution and avoided the extensive calculations of an exact permutation approach.

6.1.2 1980–1999

The advancement of permutation tests in the 1980s and 1990s was as a result of greatly improved computer clock speeds and widely available desktop computers (Good; 2005). At the same time, there was a shift in the sources of permutation publications. In the 1960s and 1970s, the majority of the all the published papers on permutation methods appeared in computer journals such as *The Computer Journal*. In the 1980s and 1990s, there was a shift away from computer journals to statistical journals - such as *The Journal of the American Statistical Association* and *Applied Statistics* (Berry et al.; 2011). An increasing number of published papers on permutation procedures began appearing in *Educational and Psychological Measurement*, *Econometrica*, *Ecology* and *Behaviour Research Methods*. Permutation tests branched out from their home in statistics to include a variety of other disciplines, most notably in psychology with articles by Berry and Mielke (1983),

pharmacology and physiology with an important article by Ludbrook (1994); ecology with articles by Anderson (1999); and econometrics with a significant article by Kennedy (1995).

In addition, many books on permutation methods, randomization tests and exact statistical methods appeared during this period. Edgington brought out the first edition of *Randomization Tests* in 1980, a second edition in 1987 and a third edition in 1995. Edgington's book was soon followed by Hubert's 1987 book on *Assignment Methods in Combinatorial Data Analysis*. Good's two books in 1994 *Permutation Tests: A practical Guide to Resampling Methods for Testing Hypotheses* and *Permutation, Parametric and Bootstrap Tests of Hypotheses*, Manly's first edition of *Randomization and Monte Carlo Methods in Biology* in 1991 followed by a second edition in 1997, and Good's third book *Resampling Methods: A Practical Guide to Data Analysis* in 1999.

Work also continued on improving the computational efficiency of permutation tests, inspired by the ease of calculations because of increases in computer speed and storage. Between 1980 and 1999 a number of algorithms were developed that substantially reduced computation time. Berry and Mielke (1985, 1987, 1995) enhanced the procedure by coupling recursive routines with the use of an arbitrary origin. A second algorithmic innovation was to recognise that only the variable part of a statistical formula needed to be computed for each permutation.

In 1980, Mehta and Patel introduced a network algorithm that effectively calculated permutation tests. Originally designed for exact tests on $2 \times k$ contingency tables, the algorithm was extended to the $r \times c$ contingency table. Extensions to multidimensional tables were provided by Mielke and Berry (1998).

Much of the contributions to the permutation literature during this period focussed on efficient means to calculate permutation versions of existing statistics. However, the advancements in computational efficiency allowed for the development of a wider variety of statistical tests, tailored to the problem under consideration. Permutation versions of existing statistics include the Wilcoxon signed ranks test by Dallal (1988) and a one-way analysis of variance test by Berry and Mielke (1983). At the same time, Mielke and his collaborators focussed their work on designing permutation tests that were not permutation versions of

existing statistics. Conventional statistical tests and measures, both parametric and nonparametric are based on squared Euclidean distances between data points, for example the two-sample t -tests, various F -tests, ordinary least-squares regression, and nonparametric tests such as the Wilcoxon rank sum test, the Kruskal-Wallis one-way analysis of variance by ranks.

6.1.3 2000–2012

Clock speeds on personal computers have increased significantly between 2000 and 2012. In 2000, the Intel Pentium processor contained 42 million transistors and ran at 1.5 GHz. In 2010, Intel released the Itanium processor, containing 4.8 GHz. To stress the progress of computing, in 1951 the Remington Rand Corporation introduced the Univac computer running at 1905 calculations per second with the storage space of 20,000 bytes of information. In 2008 the IBM Corporation supercomputer reached a sustained performance of 1 quadrillion calculations per second. In 2010 the Cray Jaguar was named the world's fastest computer performing at a speed of 1.75 quadrillion calculations per second with 360 terabytes of memory. In November 2010 China exceeded the computing speed of the Cray Jaguar by 57% with the introduction of the Tianhe-A1 super computer performing at 2.67 quadrillion calculations per second. The number one ranking for 2012 is the Sequoia supercomputer in California. The Sequoia's ability is to crunch 16.32 quadrillion calculations per second.

For a more general perspective, the personal computers available in 2012 contain processors of 3GHz with one terabyte of storage space. With the high demand for more powerful cell phones and ipads, there have been huge advancements in processing power. In August 2012 the new compact 8mm thin iphone 5 was released and its processing speed is more powerful than the average personal computer.

Increased computational efficiency has allowed for the establishment of a number of software packages for permutation tests, now widely available to computational statisticians. The most popular software packages available are SAS, SPSS, Statistica, S-plus and R. The R software package is the only free software that is available and thus it is very popular among

researchers. The University of Cambridge provides a permutation course for its first year students and the R software package is used in this course.

In addition to permutation software, the computer age of 2000-2012 had a number of books on permutation methods published. Examples include a second edition of *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypothesis* by Good in 2000; a second edition of *Permutation, Parametric and Bootstrap Tests of Hypotheses* by Good in 2000; *Permutation Methods: A Distance Function Approach* by Mielke and Berry in 2001; *Multivariate Permutation tests: With Applications in Biostatistics* by Pesarin in 2001; a third edition of *Permutation, Parametric and Bootstrap Tests of Hypotheses* by Good in 2005; a fourth edition of *Randomization Tests* by Edgington and a third edition of *Randomization, Bootstrap and Monte Carlo Methods in Biology* by Manly in 2007.

There have been many publications on permutation methods between 2000 and 2012, with steady increases each year. For example, in the search of "Permutation" on Wiley Online Library in 2000 there were 13,153 journals and 836 books available, in 2005 there were 19,709 journals and 2,155 books. In 2010 there were 30,510 journals and 4,315 books and in 2012 there are now 35,899 journals and 5,047 books. An examination of the fields of research in which articles using permutation methods were published includes computer science, biology, genetics, statistics, geology, conservation, epidemiology, ecology, public health, environmental research, geology, medicine, history, atmospheric science and public health.

7 Optimal Procedures

It is easy to justify an experiment when going from cause to effect. If a computer is available, various techniques can be used to generate an outcome. The difficulty comes into play when going in the opposite direction, from effect to cause, because more than one set of causes can be responsible for the effect (Manly; 2007). In real life and in real populations there are vast differences from subject to subject. To illustrate this variation, consider an example from Steyn et al. (2000). A firm interested in purchasing a new tyre brand. A firm with a fleet of vehicles of the same make has been using "Light Tread" tyres for their cars for some time.

The firm is satisfied with this type of tyre, but receives an offer from the manufacturers of -Hold & Clingø for tyres at a higher discount. The firm now has to choose between these two types of tyres. Figure 5.1 depicts the results of an experiment in which the first group of customers had -Light Treadø tyres fitted onto their car, while the second group of cars were fitted with -Hold & Clingø. Each of the customers then provided a subjective rating of the satisfaction of the tyres. The ratings ranged from -worsened treadø to -improved treadø on a scale from 0 to 4.

Figure 5.1. Customer satisfaction ratings for Light Tread and Hold & Cling tyres



The customers that used -Hold & Clingø seem to be more satisfied than those using -Light Treadø. Or perhaps the observed results may just be due to chance. If it is a chance effect, rather than one caused by the type of tyre, then an error has been made. An error will also be made if it is decided that there is no difference, when, in fact, -Hold & Clingø tyres are better. It is important to distinguish between these two types of errors since they have different implications. The type I error, a false positive, consists of labelling the new tyre as better. Such an action means economic loss for the firm and denial of the new tyre's benefit to the public. But a false negative, a Type II error, would mean exposing many people to a potentially dangerous tyre, which could lead to fatal car accidents.

Table 5.1 *Decision making under uncertainty*

| The Facts | | The Final Decision |
|--------------------|--|--|
| | No difference | New tyre is better |
| No difference | | Type I error: firm misses opportunity for profit; public denied access to tyre improvements |
| New tyre is better | Type II error: Customers injured; families suffer; firm sued. | |

Since variation is inherent in nature, there is bound to be occasional error when inferences are drawn from experiments and surveys, especially if chance supplies an unrepresentative sample. When a coin is tossed ten times, it is possible to get five heads and five tails. Even though the chance of getting ten tails is very low, it is possible. As the popular saying goes, the latter is less probable but not impossible.

The risk of making statistical decisions cannot be eliminated but can be contained by using the correct statistical procedures. If, for example, the probability of making a type I error does not exceed 1%, 5% or 10%, and the choice of statistical procedures is correct, a method will be provided to keep the type II error as small as possible.

The losses illustrated above will not only depend on whether the guess was right or wrong, but on how far off the mark the guesstimate is. Suppose a new type of pesticide has been produced for fruit pests and an investigation into the side effects on people eating this fruit has been done. Suppose a conclusion has been drawn and the pesticide does not cause any harm to people. But the truth is that the pesticide raises blood toxicity by a mere fraction of the normal day-to-day toxicity levels. Now, suppose that a slightly higher concentration in pesticide chemicals raises toxicity levels to potentially cause cancer. What will be the cost to the patients as well as to the company regarding law suits? The cost of the type II error will depend on the severity of the error and the nature of the losses associated with it.

Most of the work in hypothesis testing has been focussed on zero or one loss function, while estimation has focussed on losses proportional to the error squared. When making an assessment, it is important to consider whether the concern lies more with a specific decision or those that need to be sustained over time as a result of keeping to a specific decision (Good; 2005). What is most important, reducing average losses over time or avoiding one catastrophic loss?

In order to come to the best conclusion, the significance level (α) and the power (β) should to be looked at closely. α represents the probability of making a type I error and β represents the probability of a type II error. To test a hypothesis the set of possible outcomes is divided into two or more regions. The primary hypothesis is accepted when a type I error is risked when the test statistic lies in the rejection region. The primary hypothesis is rejected when a risk of a type II error occurs when the test statistic lies in the acceptance region. Additional observations may be taken if the test statistic lies in the region of indifference (Pesarin; 2001).

The ideal statistical test would have a significance level α of zero and a power β of 1. Since the real world is not idealistic, this ideal cannot be realised. In practise the significance level is fixed at the largest desirable level and a test statistic is chosen to maximize the power for a set of important alternatives.

For a fixed significance level, the power is an increasing function of the size of the effect. For a fixed effect, increasing the significance level also increases the power of the test. As previously discussed, the greater the difference between the true alternative and the decided hypothesis, the greater the loss associated with a type II error.

Fortunately, tests can be devised where the larger the discrepancy, the greater the power, and the less likely there is to be a type II error. The power can also be increased by increasing the sample sizes as they are directly related. A more powerful test reduces the cost of experimentation and minimizes the risk.

Figure 5.2 Power as a function of the alternative. Tests have the same sample size

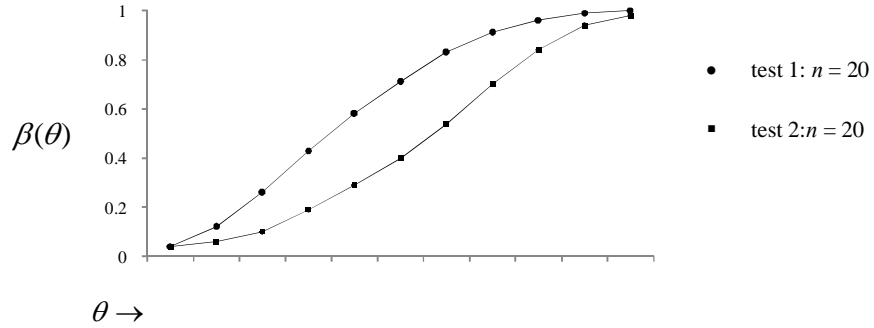


Figure 5.3 Power as a function of the alternative. Tests have different sample sizes

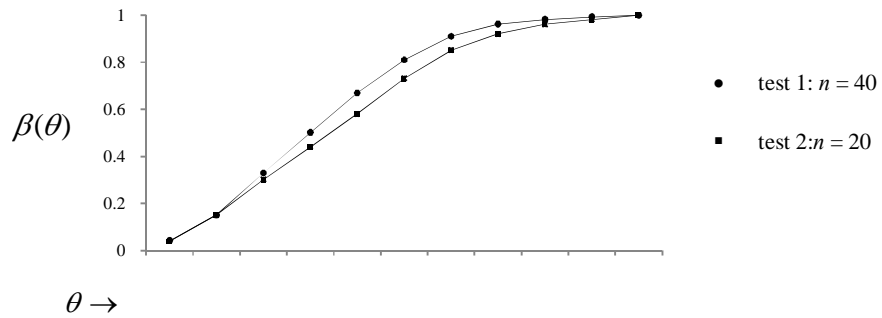
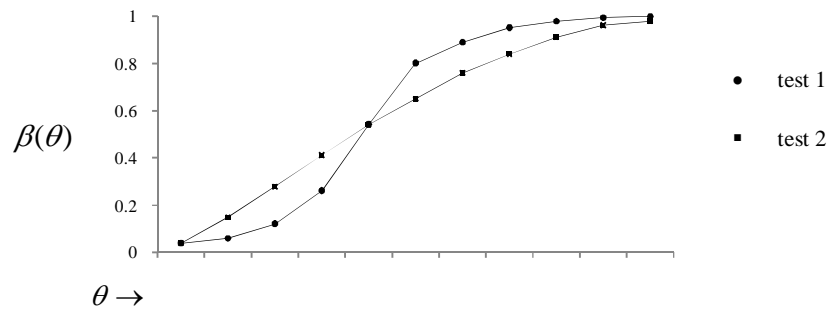


Figure 5.4 Comparing power curves. For near alternatives, with θ close to zero, test 2 is the more powerful test; for far alternatives, with θ large, test 1 is more powerful. Thus, neither test is most powerful.



Source: Good PI. *Permutation, Parametric and Bootstrap Tests of Hypotheses*. 3rd ed. New York: Springer-Verlag;2005.

The power of tests can only be compared if they have the same significance level. If the test $\omega_1(\alpha_1)$ is less powerful than $\omega_2(\alpha_2)$, where the significance level $\alpha_1 < \alpha_2$, then the power of $\omega_1(\alpha_2)$ is greater than the power of $\omega_2(\alpha_2)$. The significance level and power may also depend upon how the variables are distributed. Thus the type I error may need to be less than or equal to some predetermined value for all possible distributions. When applied correctly, permutation tests always have this property (Good; 2005). The significance levels of parametric tests and the tests based on the bootstrap technique depend on the underlying distribution. Thus permutation tests are a more powerful alternative when the underlying distribution is unknown.

8 Analysis of Variance

8.1 Introduction

Analysis of variance (ANOVA) is used to test whether various populations differ from one another in respect of a particular characteristic. In analysis of variance, the different populations are usually described as treatments. The observations on which the analysis is based are then the results that are obtained by applying these treatments to the experimental units. As with other parametric tests, the assumption of normality, independence within and between groups and homogeneity among the populations should be satisfied. If these assumptions are violated then permutation tests are a viable option.

The ANOVA method using permutation tests have been extensively studied, especially when faced with interaction terms. Several permutation strategies have been proposed to obtain a distribution-free test in ANOVA with a single error term. This chapter describes the one-way, two-way and three-way ANOVA permutation strategies that have been proposed. For the one-way ANOVA there is a general consensus regarding the permutation procedure. However, for two-way and three-way ANOVA, no universal permutation test (can be applied to an arbitrary design to test a desired factor) exists. One method is proposed by Manly (1997), which uses the permutation of the raw data. Edgington (2007) follows the methods

used by Manly (1997). Another method is that of Still and White (1981) which removes the main effects when testing for interaction. The Ter Braak (1992) method uses the full model (no effects are removed) when testing for interaction. These permutation tests are not exact procedures as there is correlation in the residuals present. An exact test is proposed by Jung et al. (2006) which removes the correlation by means of a transformation. A simulation experiment is done to compare performance of the normal F -test and the various permutation tests under conditions of normality and non-normality.

8.2 One-Way Analysis of Variance

8.2.1 The Parametric Approach

ANOVA and t -tests are used for detecting differences between experimental treatments where it is expected that some treatments will be more effective than others. The independent t -test can be regarded as a special case for where there are only two treatments. The simultaneous comparison of more than two treatment groups using the F -test can also be advantageous to a researcher. However, a significant F for a comparison of several treatments does not permit one to conclude which particular treatment differs from each other in their effects. The only justifiable statistical inference from a significant overall F is that the treatments do not have identical effects.

One-way ANOVA is not sensitive to treatment differences when many of the treatments have almost identical effects and only one or two are quite different in their effects. Whether four out of five treatments have the same effect and the fifth has a different effect cannot be determined without more specific comparisons. Follow-up tests comparing individual treatments in pairs by use of t -tests are sometimes conducted and, either one-tailed or two-tailed tests can be used for this purpose (Edgington; 2007).

For the one-way ANOVA, assume that y_{i1}, \dots, y_{in_i} is a random sample from a $N(\mu_i, \sigma^2)$ population, homogeneity between the populations is present and that the samples are independent. Although the null hypothesis of equality means should be formulated as $\mu_1 = \dots = \mu_a$, it is customary to regard μ_i as the sum of an overall mean component such as μ and a component due to the specific population. For instance, write $\mu_i = \mu + (\mu_i - \mu) = \mu + \alpha_i$ where $\alpha_i = \mu_i - \mu$. The reparameterization leads to the hypothesis of means

$$H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_a \quad \text{versus} \quad H_1 : H_0 \text{ not true}$$

The response y_{ik} , distributed as $N(\mu + \alpha_i, \sigma^2)$, can be expressed in the form

$$y_{ik} = \mu + \alpha_i + \varepsilon_{ik}, \quad i = 1, \dots, a \text{ and } k = 1, \dots, n_i$$

where the ε_{ik} are independent $N(0, \sigma^2)$ random variables. To define uniquely the model parameters and their least squares estimates, is customary to impose the constraint

$$\sum_{i=1}^a n_i \alpha_i = 0$$

The usual F -test rejects H_0 at a α significance level if

$$F = \frac{\sum_{i=1}^a n_i (\bar{y}_i - \bar{y}_{..})^2 / (a-1)}{\sum_{i=1}^a \sum_{k=1}^{n_i} (y_{ik} - \bar{y}_i)^2 / (\sum_{i=1}^a n_i - k)} > F_{a-1, \sum n_i - a}(\alpha)$$

Where $F_{a-1, \sum n_i - a}(\alpha)$ is the upper (100α) th percentile of the F -distribution with $a-1$ and

$\sum n_i - a$ degrees of freedom respectively (Johnson and Wichern; 2002).

8.2.2 The Permutation Approach

A systematic way of listing data permutations is necessary for the systematic permutation method; this ensures that all data permutations are considered. Permutations for one-way ANOVA can be listed by making use of a numerical example:

Suppose there are two measurements for treatment A, two for treatment B, and three for treatment C, as follows:

| A | B | C |
|-------|--------|------------|
| 17, 8 | 19, 25 | 24, 17, 15 |

To systematically list the $7! / 2!2!3! = 210$ permutations, index numbers one to seven are assigned to the seven measurements:

| A | B | C |
|-------|--------|------------|
| 17, 8 | 19, 25 | 24, 17, 15 |
| 1, 2 | 3, 4 | 5, 6, 7 |

The permutation of index numbers can be represented as a seven digit number: 1234567. The permutations of the index numbers are listed in order of magnitude from the smallest seven digit number, keeping the index values in ascending order within a treatment. This ensures no redundancy results from the same combination. The following listing shows the first two and last two permutations of index numbers, when listed by the described procedure:

| Permutation Number | A | B | C |
|--------------------|-----|-----|-------|
| 1 | 1 2 | 3 4 | 5 6 7 |
| 2 | 1 2 | 3 5 | 4 6 7 |
| ... | | | |
| 209 | 6 7 | 3 5 | 1 2 4 |
| 210 | 6 7 | 4 5 | 1 2 3 |

Suppose the null hypothesis $H_0 : \alpha_A = \alpha_B = \alpha_C$ is true and the groups are not really different (in terms of measured variable). If this is the case, then the observations are exchangeable between the different groups. That is, the labels that identify them as belonging to a particular group can be randomly permuted to obtain a new value of F , denoted F^* . If all possible values for F^* are calculated for all the different allocations of the labels to the observed values, this would give the entire distribution of the F statistic under a true null hypothesis, given the particular data set.

To calculate the p-value for the test, compare the value of F calculated on the original data with the distribution of values F^* obtained for a true null by permuting the labels. The empirical frequency distribution of F^* can be articulated entirely; that is, the number of possible ways that the data can be re-ordered is finite. The probability associated with the null hypothesis is calculated as the proportion of the F^* greater or equal to F .

$$\text{p-value} = \frac{\text{number of } F^* \geq F}{\text{total number of } F^*} = P(F^* \geq F)$$

In this calculation, the observed value is included as a member of the distribution. This is because one of the possible random orderings of the treatment labels is the ordering that was actually obtained. This p-value gives an exact test of the null hypothesis of no differences among groups (Hoeffding; 1952).

An equivalent test statistic described by Pesarin (2001) for the one-way ANOVA is

$$T = \sum_{i=1}^a n_i \bar{y}_i^2 \quad \text{where } \bar{y}_i = \sum_k^n y_{ik} / n_i.$$

This statistic is equivalent to $\sum_i^a n_i (\bar{y}_i - \bar{y}_{..})^2$, where $y_{..} = \sum_{ik}^{an} y_{ik} / n$. Since $(a-1) / \sum n_i - a$ is a constant multiplier over all permutations, its elimination has no effect on the ordering of the permutations with respect to the test statistic value. A requirement is that the errors ε_{ik}

should be exchangeable with respect to groups only in H_0 . To show that T is an exact permutation test, consider its permutation structure and assume that v_{hi}^* data are randomly moved from the h^{th} to the i^{th} group, where $\sum_h v_{hi}^* = n_i$ for $i=1, \dots, a$, and where v_{hh}^* represents the number of observations which remain in the h^{th} group. The permutation structure of the statistic

$$T^* = \frac{\sum_i^a \left(\sum_k^n (\alpha_i^* + \varepsilon_{ik}^*) \right)^2}{n_i}$$

shows that, if and only if H_0 is true, T^* depends only on a permutation of exchangeable errors, whereas H_1 depends on treatment effects as well. Hence, as the permutation null distribution of T depends only on exchangeable errors, T is an exact permutation test. The computation of T as a test statistic will give the same p-value as the computation of F and is thus easier than the F -statistic to calculate.

8.3 Two-Way Analysis of Variance

8.3.1 The Parametric Approach

Taking an example from Manly (2007), the concept of interactions can be illustrated. Suppose we are concerned with the number of ants consumed by two sizes of lizards over each of the four months given below.

Table 8.3.1 The number of ants eaten from June to September by small and large lizards.

| The Facts | Small | | | Large | | |
|------------------|-------|-----|-----|-------|------|-----|
| <i>June</i> | 13 | 242 | 105 | 182 | 21 | 7 |
| <i>July</i> | 8 | 59 | 20 | 24 | 312 | 68 |
| <i>August</i> | 515 | 488 | 88 | 460 | 1223 | 990 |
| <i>September</i> | 18 | 44 | 21 | 140 | 40 | 27 |

It appears as if there is the possibility of an interaction since in June the small lizards ate more ants than the large lizards, but the reverse happened in the other months. There might also be an effect for size, with large lizards eating more ants than small lizards. Looking at an analysis of variance summary table the assumptions can be confirmed. The interaction is border line, as is the size effect, yet the months effect is significant.

Table 8.3.2 Output from a standard ANOVA for size and months on the number of ants consumed

The ANOVA Procedure

Dependent Variable: Ants_Consumed

| Source | DF | Anova SS | Mean Square | F Value | Pr > F |
|--------------------|----|-------------|-------------|---------|--------|
| <i>size</i> | 1 | 146172.042 | 146172.042 | 4.47 | 0.0505 |
| <i>months</i> | 3 | 1379495.125 | 459831.708 | 14.06 | <.0001 |
| <i>size*months</i> | 3 | 294009.458 | 98003.153 | 3.00 | 0.0617 |

In general, suppose we want to assess the simultaneous effects on number of ants eaten, given the month and size of the lizard. Let the number of ants eaten be represented by y_{ijk} for a different month and b different sizes of lizards, where $i = 1, \dots, a$; $j = 1, \dots, b$, and n represents observations at each factor combination $i j$, $k = 1, \dots, n$. A model relating the dependent variable, ants eaten (the effect) to the independent variables of months and size (the causes) is given by

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk} \quad \text{or}$$

$$y_{ijk} = \mu + \text{Size}_i + \text{Month}_j + \text{Interaction}_{ij} + \varepsilon_{ijk}$$

Under the normality assumption of error terms, the traditional F -test statistic is

$$F = \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2 / (a-1)(b-1)}{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (y_{ijk} - \bar{y}_{ij.})^2 / ab(n-1)}$$

where $\bar{y}_{ij.} = (\sum_{k=1}^n y_{ijk}) / n$, $\bar{y}_{i..} = (\sum_{j=1}^b \sum_{k=1}^n y_{ijk}) / bn$, $\bar{y}_{.j.} = (\sum_{i=1}^a \sum_{k=1}^n y_{ijk}) / an$ and $\bar{y}_{...} = (\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n y_{ijk}) / abn$.

The null and alternative hypothesis for the interactions of the model are

$$H_0 : (\alpha\beta)_{11} = \dots = (\alpha\beta)_{ab} = 0 \quad \text{versus} \quad H_1 : \text{not } H_0 .$$

Under the null hypothesis of no interaction, F is distributed as $F((a-1)(b-1), ab(n-1))$ (Johnson and Wichern; 2002).

8.3.2 Permutation of Raw Data

8.3.2.1 The Manly Method

Manly (2007) suggests that one way to permute the data is to randomize the samples over all the cells in the experiment. For example, take the values, shake them up in a bowl, and write them down in whatever order they come out of the bowl next to the columns that contain information on the variables. Considering the previous example on lizards, these variables would be size and months. The F -values for the effects is calculated, these values are then stored away, and then this procedure is repeated another 4,999 times⁽¹⁾. At the end there would be 5,000 values of F^M that would reasonably occur under the null hypothesis. The original F -statistic obtained can then be compared against the empirical under the null hypothesis, that is, $H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_a$. The p-value is then calculated as the number of resampled F^{M*} values that exceed the original F , that is,

$$\text{p-value} = P(F^{M*} \geq F) = \frac{\text{number of } F^{M*} \geq F}{n!}$$

Let this test is denoted by PT1. The same procedure is followed for the two main effects. This procedure by Manly (1997) is perhaps the easiest to carry out.

8.3.2.2 *The Edgington Method*

Edgington (2007) goes at the problem from a different direction. He maintains that there is no exact test for interactions, but suggests that one can get an indication of the presence of interactions by testing the interaction in the same way that Manly (1997) does. However, only the interaction is tested in this way. Edgington (2007) permutes all observations 5,000 times across cells, computes the F -values for the interaction for each one, and then calculates $P(F^{E*} \geq F)$.

According to Edgington (2007), there are two ways to deal with the main effects. If the interaction is significant, one probably doesn't care about the main effects. It is alright to look at the main effects if one has good reason, but one seldom has a good reason to want to deal with the main effects when faced with an interaction.

If the interaction effect is not significant, one would probably want to go ahead and deal with main effects. One way to do this is the same way Manly (1997) does. Edgington(2007) would argue that if there is no interaction, the best model is $y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk}$ or $y_{ijk} = \mu + Size_i + Month_j + \varepsilon_{ijk}$ which is an additive model ó it does not have an interaction. Edgington (2007) reasons that if this is now the model, one does not have to adjust for an interaction. Therefore one can test the size by shuffling the data for each month separately between size categories, and then test months by shuffling the data for each size among the month categories. In each of these steps the distribution of the randomised F -values are formed. The original F -value is then compared against the permuted F -values.

8.3.3 Permutation under the Reduced Model

8.3.3.1 The Still and White Method

Still and White (1981) follow Edgington (2007) in using restricted randomization for main effects, but they control for main effects by using residuals. In this method, the residuals are computed from a model that includes all parameters except the parameters of interest. This means that only the effect of the parameters that are of interest are possibly present in these residuals. In this case, the residuals are what are left over after the row and column effects are removed. If one is looking at all the data combined, there are potential main effects of size and months included in them. But if these effects are subtracted and the residuals are used, a test for interaction can be performed, without worrying about the main effects. Since any possible row and column effects have been removed, the residuals are exchangeable under $H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_a$. One simple way to do this is to compute

$$y_{ijk}^{SW} = y_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}$$

However, under the null hypothesis, there should be no systematic effect in these residuals and therefore an unrestricted permutation of these residuals is used to test the parameters of interest. If a complete analysis of variance is run on the residuals, the sums of squares for the effects are zero, and the F for the interaction is a fair test of that interaction uncontaminated by the main effects. Anderson and Ter Braak (1992) show that this method has relatively more power in comparison with the other analysis of variance methods.

Using the Still and White (1981) approach on randomization, an F -statistic can be obtained by the following:

$$F^{SW} = \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{y}_{ij.}^{SW} - \bar{y}_{i..}^{SW} - \bar{y}_{.j.}^{SW} + \bar{y}_{...}^{SW})^2 / (a-1)(b-1)}{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{y}_{ijk}^{SW} - \bar{y}_{ij.}^{SW})^2 / ab(n-1)}$$

$F_{AB}^{SW} = F_{AB}$ since it is obtained directly from the raw data. Hence the statistical significance of H_0 can be evaluated from the Monte Carlo distribution of

$$F_{AB}^{SW*} = \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{y}_{ij.}^{SW*} - \bar{y}_{i..}^{SW*} - \bar{y}_{.j.}^{SW*} + \bar{y}_{...}^{SW*})^2 / (a-1)(b-1)}{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{y}_{ijk}^{SW*} - \bar{y}_{ij.}^{SW*})^2 / ab(n-1)}$$

Where y_{ijk}^{SW*} is an ijk^{th} element of \mathbf{y}^{SW*} which is an $abn \times 1$ randomly permuted vector of $\mathbf{y}^{SW} = (y_{111}^{SW}, \dots, y_{11n}^{SW}, \dots, y_{abn}^{SW})$. Thus the significance of H_0 can be assessed by

$$\text{p-value} = P(F^{SW*} \geq F) = \frac{\text{number of } F^{SW*} \geq F}{n!}$$

Let this test be denoted by PT2. This test however is based on the correlated residuals and it will not be exact for finite samples. For example, with the residuals $y_{ijk}^{SW} = y_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}$, two residuals in the same row, such as y_{i11}^{SW} and y_{i23}^{SW} will be correlated, while two observations from different rows and columns will not. The residuals are not exchangeable, thus the distribution with respect to all possible permutations will not be exact. However, following Good (2002), the studentized correlations between the residuals converge to a common value as the sample size increases; thus the residuals are asymptotically exchangeable, and this test is asymptotically exact.

8.3.4 Permutation of the Full Model

8.3.4.1 The Ter Braak Method

Ter Braak (1992) has done a great deal of work with randomization procedures and he advocates an approach similar to the Still and White (1981) approach, except that in calculating the interaction, the residuals are taken over the whole design rather than just the additive model. The cell mean is subtracted so that the effects of all factors are removed from the original observations. A new sample is then constructed by adding the fitted value to an unrestricted permuted version of the residuals. This approach is very much like the Still and

White (1981) approach, but only row and column means are subtracted, not individual cell means. Again this amounts to fitting a complete model (including the interaction) and computing the residuals. Ter Braak (1992) uses the usual residuals $y_{ijk}^{TB} = y_{ijk} - \bar{y}_{ij}$ instead of $y_{ijk}^{SW} = y_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}$. By using the freely randomized residuals y_{ijk}^{TB} , F_{AB}^{TB} can be calculated from the normal F -statistic. Then the significance of $H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_a$ can be assessed by

$$\text{p-value} = P(F^{TB*} \geq F) = \frac{\text{number of } F^{TB*} \geq F}{n!}$$

Let this test be denoted by PT3. It will however not be exact for a finite sample since this test is based on the correlated residuals $y_{ijk}^{TB} = y_{ijk} - \bar{y}_{ij}$.

8.3.5 Permutation for an Exact Test

8.3.5.1 The Jung Method

The permutation tests described by Manly (1997), Edgington (2007) and Still and White (1981) are not exact tests, since the probability distribution of the permuted observations is different from that of the original observations, even when no interaction effect exists. An exact random permutation test is proposed by Jung et al. (2006), which is an improvement to the Still and White (1981) approach. This is based on the uncorrelated residuals obtained from a transformation of the correlated residuals $y_{ijk}^{SW} = y_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}$. To obtain a new random permutation test, the model

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk},$$

can be written in matrix form as

$$\mathbf{y} = \mathbf{X}_0\boldsymbol{\mu} + \mathbf{X}_a\boldsymbol{\alpha} + \mathbf{X}_b\boldsymbol{\beta} + \mathbf{X}_{ab}\boldsymbol{\alpha\beta} + \boldsymbol{\varepsilon},$$

where $\mathbf{y} = (y_{111}, \dots, y_{11n}, \dots, y_{abn})'$ is a $abn \times 1$ response vector, and $\boldsymbol{\varepsilon} = (\varepsilon_{111}, \dots, \varepsilon_{11n}, \dots, \varepsilon_{abn})'$ is a $abn \times 1$ error vector, $\mathbf{X}_0 = \mathbf{i}_a \otimes \mathbf{i}_b \otimes \mathbf{i}_K$, $\mathbf{X}_a = \mathbf{I}_a \otimes \mathbf{i}_b \otimes \mathbf{i}_K$, $\mathbf{X}_b = \mathbf{i}_a \otimes \mathbf{I}_b \otimes \mathbf{i}_K$, where $\mathbf{i}_a, \mathbf{i}_b$ and \mathbf{i}_n are vectors of ones of dimension a, b and n , respectively and \mathbf{I}_a and \mathbf{I}_b are identity matrices of dimension a and b respectively; \otimes denotes the kronecker product, and $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_a)'$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_b)'$, and $\boldsymbol{\alpha}\boldsymbol{\beta} = ((\boldsymbol{\alpha}\boldsymbol{\beta})_{11}, \dots, (\boldsymbol{\alpha}\boldsymbol{\beta})_{1b}, \dots, (\boldsymbol{\alpha}\boldsymbol{\beta})_{ab})'$.

Define $\mathbf{X}_{ab} = (\mathbf{X}_a : \mathbf{X}_b)$ and $\mathbf{H}_{ab} = \mathbf{X}_{ab}(\mathbf{X}_{ab}'\mathbf{X}_{ab})^{-1}\mathbf{X}_{ab}'$, then it can be shown that $\mathbf{H}_{ab} = \mathbf{I}_a \otimes \mathbf{J}_b \otimes \mathbf{J}_n + \mathbf{J}_a \otimes \mathbf{I}_b \otimes \mathbf{J}_n - \mathbf{J}_a \otimes \mathbf{J}_b \otimes \mathbf{J}_n$ using the generalized inverse of the partitioned matrix, where $\mathbf{J}_a = \mathbf{i}_a \mathbf{i}_a' / a$, $\mathbf{J}_b = \mathbf{i}_b \mathbf{i}_b' / b$, and $\mathbf{J}_n = \mathbf{i}_n \mathbf{i}_n' / n$ are the matrix of dimension a, b and n whose elements are all $1/a, 1/b$, and $1/n$, respectively. Then, multiplying $\mathbf{I}_{abn} - \mathbf{H}_{ab}$ to both sides of equation $\mathbf{y} = \mathbf{X}_0\boldsymbol{\mu} + \mathbf{X}_a\boldsymbol{\alpha} + \mathbf{X}_b\boldsymbol{\beta} + \mathbf{X}_{ab}\boldsymbol{\alpha}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ one gets

$$\mathbf{y}^{J1} = \mathbf{X}_{ab}^{J1}\boldsymbol{\alpha}\boldsymbol{\beta} + \boldsymbol{\varepsilon}^{J1},$$

with

$$\mathbf{y}^{J1} = (\mathbf{I}_{abn} - \mathbf{H}_{ab})\mathbf{y}, \quad \mathbf{X}_{ab}^{J1} = (\mathbf{I}_{abn} - \mathbf{H}_{ab})\mathbf{X}_{ab}, \quad \boldsymbol{\varepsilon}^{J1} = (\mathbf{I}_{abn} - \mathbf{H}_{ab})\boldsymbol{\varepsilon}.$$

where the ijk^{th} element of \mathbf{y}^{J1} is $y_{ijk}^{J1} = y_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}$. The idea is to remove part of the design matrix corresponding to the parameters not of interest, that is, the main effects. The traditional F -test statistic for the interaction is derived for the transformed model as

$$F^{J1} = \frac{\mathbf{y}^{J1}'(\mathbf{X}_{ab}^{J1}(\mathbf{X}_{ab}^{J1}\mathbf{X}_{ab}^{J1})^{-1}\mathbf{X}_{ab}^{J1})\mathbf{y}^{J1}/(a-1)(b-1)}{\mathbf{y}^{J1}'(\mathbf{I}_{abn} - \mathbf{X}_{ab}^{J1}(\mathbf{X}_{ab}^{J1}\mathbf{X}_{ab}^{J1})^{-1}\mathbf{X}_{ab}^{J1})\mathbf{y}^{J1}/ab(n-1)}$$

Thus the permutation method using \mathbf{y}^{J1*} a randomly permuted vector of \mathbf{y}^{J1} is the same method as that of Still and White (1981). For the permutation method of Still and White (1981), it can be noted that

$$\boldsymbol{\varepsilon}^{J1} \sim (\mathbf{0}, \sigma^2(\mathbf{I}_{abn} - \mathbf{H}_{ab}))$$

Consequently, $\boldsymbol{\varepsilon}^{J1}$, a randomly permuted vector of $\boldsymbol{\varepsilon}^{J1}$ is distributed differently from $\boldsymbol{\varepsilon}^{J1}$. Hence, even though under $H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_a$, the distribution of \mathbf{y}^{J1*} is not equal to that of \mathbf{y}^{J1} . This fact may damage the rationale behind the random permutation test. To resolve this problem Jung et al. (2006) proposes a transformation using the decomposition of the idempotent matrix $\mathbf{I}_{abn} - \mathbf{H}_{ab}$. Since $\mathbf{I}_{abn} - \mathbf{H}_{ab}$ is a symmetric and idempotent matrix, it possesses two distinct eigenvalues, zero and one. Let $\mathbf{I}_{abn} - \mathbf{H}_{ab} = \mathbf{UDU}'$ be the eigen-decomposition of $\mathbf{I}_{abn} - \mathbf{H}_{ab}$, where \mathbf{D} is a diagonal matrix containing the eigenvalues of $\mathbf{I}_{abn} - \mathbf{H}_{ab}$, and \mathbf{U} is a unitary matrix, whose columns are the eigenvectors of $\mathbf{I}_{abn} - \mathbf{H}_{ab}$.

Since $\mathbf{I}_{abn} - \mathbf{H}_{ab}$ is a symmetric and idempotent matrix of rank $abn - a - b + 1$, there exists a matrix \mathbf{V} of dimension $abn \times (abn - a - b + 1)$. Orthonormality of the columns of \mathbf{V} implies that $\mathbf{V}'\mathbf{V} = \mathbf{I}_{abn-a-b+1}$ where

$$\begin{pmatrix} v_{1,1} & \cdots & v_{1,abn-a-b+1} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ v_{abn,1} & \cdots & v_{abn,abn-a-b+1} \end{pmatrix} \begin{pmatrix} v_{1,1} & \cdots & v_{1,abn} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ v_{abn-a-b+1,1} & \cdots & v_{abn-a-b+1,abn} \end{pmatrix} = \begin{pmatrix} 1 & \cdots & 0 \\ \cdot & 1 & \cdot \\ \cdot & & \cdot \\ \cdot & & 1 \cdot \\ 0 & \cdots & 1 \end{pmatrix}$$

since $v_{ik} = \begin{cases} 1 & \text{for } k=i \\ 0 & \text{for } k \neq i \end{cases}$, thus the sum of the corresponding projections are equal to the identity matrix. Since

$$\mathbf{I}_{abn} = \mathbf{X}_{ab}(\mathbf{X}_{ab}'\mathbf{X}_{ab})^{-1}\mathbf{X}_{ab}' + \mathbf{V}(\mathbf{V}'\mathbf{V})^{-1}\mathbf{V}'$$

one gets,

$$\begin{aligned} \mathbf{I}_{abn} &= \mathbf{X}_{ab}(\mathbf{X}_{ab}'\mathbf{X}_{ab})^{-1}\mathbf{X}_{ab}' + \mathbf{V}(\mathbf{I}_{abn-a-b+1})^{-1}\mathbf{V}' \\ \mathbf{I}_{abn} &= \mathbf{X}_{ab}(\mathbf{X}_{ab}'\mathbf{X}_{ab})^{-1}\mathbf{X}_{ab}' + \mathbf{V}\mathbf{V}' \\ \mathbf{I}_{abn} &= \mathbf{H}_{ab} + \mathbf{V}\mathbf{V}' \\ \mathbf{V}\mathbf{V}' &= \mathbf{I}_{abn} - \mathbf{H}_{ab} \end{aligned}$$

Thus the matrix \mathbf{V} satisfies the following equations:

$$\mathbf{V}\mathbf{V}' = \mathbf{I}_{abn} - \mathbf{H}_{ab} \quad \mathbf{V}'\mathbf{V} = \mathbf{I}_{abn-a-b+1}$$

Multiplying \mathbf{V}' to both sides of $\mathbf{y}^{J1} = \mathbf{X}_{ab}^{J1}\boldsymbol{\alpha}\boldsymbol{\beta} + \boldsymbol{\varepsilon}^{J1}$, one gets

$$\mathbf{y}^{J2} = \mathbf{X}_{ab}^{J2}\boldsymbol{\alpha}\boldsymbol{\beta} + \boldsymbol{\varepsilon}^{J2} \quad (8.3.1)$$

where

$$\mathbf{y}^{J2} = \mathbf{V}'\mathbf{y}^{J1}, \mathbf{X}_{ab}^{J2} = \mathbf{V}'\mathbf{X}_{ab}^{J1}, \boldsymbol{\varepsilon}^{J2} = \mathbf{V}'\boldsymbol{\varepsilon}^{J1}$$

Since $\boldsymbol{\varepsilon}^{J2} \sim (0, \sigma^2\mathbf{I}_{abn-a-b+1})$ and $\boldsymbol{\varepsilon}^{J2*}$ a randomly permuted vector of $\boldsymbol{\varepsilon}^{J2}$, one gets

$$\boldsymbol{\varepsilon}^{J2*} \sim (0, \sigma^2\mathbf{I}_{abn-a-b+1})$$

and

$$\text{var}(\boldsymbol{\varepsilon}^{J2}) = \text{var}(\mathbf{V}'\boldsymbol{\varepsilon}^{J1}) = \mathbf{V}'\text{var}(\boldsymbol{\varepsilon}^{J1})\mathbf{V} = \mathbf{V}'(\sigma^2(\mathbf{I} - \mathbf{H}_1))\mathbf{V} = \sigma^2\mathbf{V}'\mathbf{V}\mathbf{V}'\mathbf{V} = \sigma^2\mathbf{I}_{abn-a-b+1}$$

Therefore the F -statistic, F^{J2} can be obtained for testing the hypothesis of interaction, where

$$F^{J2} = \frac{\mathbf{y}^{J2}'(\mathbf{X}_{ab}^{J2}(\mathbf{X}_{ab}^{J2}'\mathbf{X}_{ab}^{J2})^{-1}\mathbf{X}_{ab}^{J2}')\mathbf{y}^{J2} / (a-1)(b-1)}{\mathbf{y}^{J2}'(\mathbf{I}_{abn-a-b+1} - \mathbf{X}_{ab}^{J2}(\mathbf{X}_{ab}^{J2}'\mathbf{X}_{ab}^{J2})^{-1}\mathbf{X}_{ab}^{J2}')\mathbf{y}^{J2} / (ab(n-1))} \quad (8.3.2)$$

Since

$$\begin{aligned} \mathbf{y}^{J2} \prime \mathbf{X}_{ab}^{J2} &= \mathbf{y}^{J1} \prime \mathbf{V} \mathbf{V} \prime \mathbf{X}_B^{J1} \\ &= \mathbf{y}^{J1} \prime (\mathbf{I}_{abn} - \mathbf{H}_{ab}) \mathbf{X}_{ab}^{J1} \\ &= \mathbf{y}^{J1} \prime \mathbf{X}_B^{J1} \end{aligned}$$

and

$$\begin{aligned} \mathbf{X}_{ab}^{J2} \prime \mathbf{X}_{ab}^{J2} &= \mathbf{X}_{ab}^{J1} \prime \mathbf{V} \mathbf{V} \prime \mathbf{X}_{ab}^{J1} \\ &= \mathbf{X}_{ab}^{J1} \prime (\mathbf{I}_{abn} - \mathbf{H}_{ab}) \mathbf{X}_{ab}^{J1} \\ &= \mathbf{X}_{ab}^{J1} \prime \mathbf{X}_{ab}^{J1} \end{aligned}$$

and

$$\begin{aligned} \mathbf{y}^{J2} \prime \mathbf{y}^{J2} &= \mathbf{y}^{J1} \prime \mathbf{V} \mathbf{V} \prime \mathbf{y}^{J1} \\ &= \mathbf{y}^{J1} \prime (\mathbf{I}_{abn} - \mathbf{H}_{ab}) \mathbf{y}^{J1} \\ &= \mathbf{y}^{J1} \prime \mathbf{y}^{J1} \end{aligned}$$

The equality $F_{AB}^{J2} = F_{AB}^{J1} = F_{AB}$ is obtained.

Under $H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_a$, \mathbf{y}^{J2} is not related to \mathbf{X}_{AB}^{J2} in (8.3.2), thus the significance of the null hypothesis can be evaluated from the Monte Carlo distribution of

$$F^{J2*} = \frac{\mathbf{y}^{J2*} \prime (\mathbf{X}_{ab}^{J2} (\mathbf{X}_{ab}^{J2} \prime \mathbf{X}_{ab}^{J2})^{-1} \mathbf{X}_{ab}^{J2} \prime) \mathbf{y}^{J2*} / (a-1)(b-1)}{\mathbf{y}^{J2*} \prime (\mathbf{I}_{abn-a-b+1} - \mathbf{X}_{ab}^{J2} (\mathbf{X}_{ab}^{J2} \prime \mathbf{X}_{ab}^{J2})^{-1} \mathbf{X}_{ab}^{J2} \prime) \mathbf{y}^{J2*} / (ab(n-1))}$$

where \mathbf{y}^{J2*} is an $(abn - a - b + 1) \times 1$ randomly permuted vector of \mathbf{y}^{J2} . Thus the significance level of H_0 can be assessed by

$$\text{p-value} = P(F^{J2*} \geq F) = \frac{\text{number of } F^{J2*} \geq F}{n!} \leq \alpha$$

The PT4 test can be explained as follows. From model $\mathbf{y}^{J1} = \mathbf{X}_{ab}^{J1} \boldsymbol{\alpha} \boldsymbol{\beta} + \boldsymbol{\varepsilon}^{J1}$, since the residuals are correlated with each other, uncorrelated residuals are obtained using the transformation of the uncorrelated residuals. A transformation matrix using the decomposition of the idempotent matrix is used. The proposed randomization is based on these uncorrelated residuals. Since the probability distribution of the permuted observations \mathbf{y}^{J2*} is the same as \mathbf{y}^{J2} , PT4 satisfies the exchangeability property and is an exact permutation test.

8.4 Three-Way Analysis of Variance

Based on invariance and sufficiency, Welch (1990) considered testing for interaction effects in the three-way ANOVA model

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\gamma)_{ij} + (\beta\gamma)_{jk} + \varepsilon_{ijk} \quad (8.4.1)$$

$$\text{for } i=1, \dots, a, \quad j=1, \dots, b, \quad k=1, \dots, c$$

with proper constraints for the parameters. In equation (8.4.1) the hypotheses for testing the interactions between factors B and C are the following:

$$H_0 : (\beta\gamma)_{11} = \dots = (\beta\gamma)_{bc} = 0 \quad \text{versus} \quad H_1 : \text{not } H_0. \quad (8.4.2)$$

Under the normality assumption of error terms, the traditional F -test is

$$F = \frac{\frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c (\bar{y}_{.jk} - \bar{y}_{.j.} - \bar{y}_{.k.} + \bar{y}_{...})^2}{(b-1)(c-1)}}{\frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c (\bar{y}_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} - \bar{y}_{.k.} + \bar{y}_{ij.} + \bar{y}_{i.k} + \bar{y}_{.jk} - \bar{y}_{...})^2}{(a-1)(b-1)(c-1)}}.$$

Under the null hypothesis of (8.4.2), F is distributed as $F((b-1)(c-1), (a-1)(b-1)(c-1))$.

The permutation methods described in the two-way analysis of variance can be extended to this problem. In this model the adjusted values used in permutations are y_{ijk}^M for the Manly (2007) method, $y_{ijk}^{SW} = y_{ijk} - \bar{y}_{ij.} - \bar{y}_{i.k} + \bar{y}_{i..}$ for the Still and White (1981) method and $y_{ijk}^{TB} = y_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} - \bar{y}_{.k} + \bar{y}_{ij.} + \bar{y}_{i.k} + \bar{y}_{.jk} - \bar{y}_{...}$ for the Ter Braak (1992) method. The Jung et al. (2006) method is similar for this model is similar to that of the two-way ANOVA model. Since $\varepsilon_{ijk}^{J1} = \varepsilon_{ijk} - \bar{\varepsilon}_{ij.} - \bar{\varepsilon}_{i.k} + \bar{\varepsilon}_{i..}$ are correlated, the uncorrelated error term ε_{ijk}^{J2} can be obtained using the singular transformation used in (8.3.1). Thus the distribution of F^{J2*} can be obtained using the random permutation of uncorrelated observations y_{ijk}^{J2*} . The p-value of PT4 can be calculated by comparing the original F -value to the distribution of F^{J2*} .

8.5 Simulation Study

To explore the performance of the procedures previously discussed for testing the hypothesis $H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_a$ in the two-way ANOVA model, a simulation study similar to that of Jung et al. (2006) is performed. The main effects $\alpha_i, i = 1, \dots, a$ and $\beta_j, j = 1, \dots, b$ are generated from uniform (-50, 50) and uniform (-20, 20) distributions respectively with constraints $\sum_{i=1}^a \alpha_i = 0$ and $\sum_{j=1}^b \beta_j = 0$. To obtain estimates for the significance level, $(\alpha\beta)_{ij}$ equals zero for all $i = 1, \dots, a$ and $j = 1, \dots, b$. For these generated values, the error term ε_{ijk} is generated from four different distributions, the $N(0,1)$, $\exp(1)-1$, $t(4)$ and $U(-4,4)$. The permutation tests are based on 1,000 replications and the above procedure is repeated 10,000 times independently in order to estimate the significance level. The normal F -test and the four permutation tests: the Manly (1997), Still and White (1981), Ter Braak (1992) and Jung et al. (2006) methods are applied at a significance level of $\alpha = 0.05$.

For $(a,b,n) = (2, 2, 2)$, when the error term is normal, the traditional F -test performs well as expected. However, when the error term follows a non-normal distribution such as the $\exp(1)-1$, $t(4)$ and $U(-\sqrt{3}, \sqrt{3})$ distribution, the significance level of the F -test is underestimated. The Still and White (1981) method underestimates the nominal significance level, while the Ter Braak (1992) and Manly (1997) methods overestimate it for all

considered distributions of the error terms. The reason for these discrepancies is probably due to the fact that the Still and White (1981), Ter Braak (1992) and Manly (1997) methods are approximate tests as the randomly permuted observations are correlated. For normal and non-normal errors, the estimated significance level not significantly different from 0.05.

As a , b and n increases, the Manly (1997), Still and White (1981) and Ter Braak (1992) methods improve. This is because the correlations between the randomly permuted observations get weaker. These results are coincide to those obtained by Jung et al. (2006) where simulations show that the difference between the procedures gets smaller as a , b and n increases.

Table 8.5.1 Average significance level for various (a,b,n) combinations at nominal size $\alpha = 0.05$ with normal, exponential, t and uniform distributions of errors based on 10,000 Monte Carlo simulations.

| (a,b,n) | Tests | $N(0,1)$ | $\exp(1)-1$ | $t(4)$ | $U(-\sqrt{3},\sqrt{3})$ |
|-----------|--------------------------|----------|-------------|--------|-------------------------|
| $(2,2,2)$ | F -Test | 0.0491 | 0.0472 | 0.0404 | 0.0431 |
| | PT1 (Manly Test) | 0.0523 | 0.0535 | 0.0522 | 0.0513 |
| | PT2 (Still and White) | 0.0394 | 0.0357 | 0.0402 | 0.0409 |
| | PT3 (Ter Braak) | 0.0535 | 0.0519 | 0.0528 | 0.0605 |
| | PT4 (Jung et al.) | 0.0504 | 0.0488 | 0.0569 | 0.0531 |
| $(2,2,3)$ | F -Test | 0.0509 | 0.0435 | 0.0398 | 0.0528 |
| | PT1 (Manly Test) | 0.0439 | 0.0473 | 0.0438 | 0.0447 |
| | RT2 (Still and White) | 0.0486 | 0.0432 | 0.0456 | 0.0502 |
| | RT3 (Ter Braak) | 0.0511 | 0.0486 | 0.0469 | 0.0508 |
| | PT4 (Jung et al.) | 0.0516 | 0.0457 | 0.0466 | 0.0531 |
| $(2,3,2)$ | F -Test | 0.0482 | 0.0437 | 0.0412 | 0.0547 |
| | PT1 (Manly Test) | 0.0464 | 0.0474 | 0.0486 | 0.0490 |
| | RT2 (Still and White) | 0.0462 | 0.0497 | 0.0436 | 0.0513 |
| | RT3 | 0.0510 | 0.0533 | 0.0458 | 0.0514 |

| | | | | | |
|----------------|---------------------------------|--------|--------|--------|--------|
| | <i>(Ter Braak)</i> | | | | |
| | PT4 <i>(Jung et al.)</i> | 0.0504 | 0.0508 | 0.0484 | 0.0532 |
| <i>(2,3,3)</i> | <i>F-Test</i> | 0.0491 | 0.0423 | 0.0421 | 0.0482 |
| | PT1 <i>(Manly Test)</i> | 0.0468 | 0.0452 | 0.0442 | 0.0471 |
| | RT2 <i>(Still and White)</i> | 0.0489 | 0.0483 | 0.0491 | 0.0511 |
| | RT3 <i>(Ter Braak)</i> | 0.0514 | 0.0476 | 0.0487 | 0.0515 |
| | PT4 <i>(Jung et al.)</i> | 0.0504 | 0.0522 | 0.0459 | 0.0521 |

8.6 Numerical Example

Referring back to the example from Manly (1997) which displays the number of ants consumed by two sizes of lizards over each of the four months, the various permutation procedures are performed to test for interaction.

Table 8.6.1 *The number of ants eaten from June to September by small and large lizards*

| The Facts | Small | | | Large | | |
|-----------|-------|-----|-----|-------|------|-----|
| June | 13 | 242 | 105 | 182 | 21 | 7 |
| July | 8 | 59 | 20 | 24 | 312 | 68 |
| August | 515 | 488 | 88 | 460 | 1223 | 990 |
| September | 18 | 44 | 21 | 140 | 40 | 27 |

From table 8.6.1 it can be seen that under the normality assumption of errors, the traditional *F*-test for interaction effects proves to be significant at the 10 percent level. Thus, it is concluded that a mild degree of interaction exists between size and months. However the normality assumption is not satisfied and therefore the traditional *F*-test cannot be satisfied. The permutation tests of Manly (1997), Still and White (1981), Ter Braak (1992) and Jung et al. (2006) are applied. These permutation tests are based on 10,000 Monte Carlo replications. From the permutation p-values, all the permutation tests in this experiment are still significant

at a 10 percent level. Thus we conclude without a doubt that interaction exists between size and months.

Table 8.6.2. *The results from a number of testing procedures for the interaction effects of size and months for small and large lizards.*

| | p-value Interaction (Size*Months) |
|---------------------------------|--------------------------------------|
| <i>F-Test</i> | 0.0617 |
| PT1 <i>(Manly Test)</i> | 0.0575 |
| PT2 <i>(Still and White)</i> | 0.0602 |
| PT3 <i>(Ter Braak)</i> | 0.0578 |
| PT4 <i>(Jung et al.)</i> | 0.0593 |

9 *Multiple Linear Regression*

9.1 *Introduction*

Several different methods of permutation have been proposed to test the significance of one or more regression coefficients in a multiple linear regression model. These tests consist of approximate and exact permutation tests. Exact permutation methods have only recently been introduced with the advancement in processing speed of computers. There are essentially three different approaches for an approximate test: Unrestricted permutation of raw data, and permutation of residuals under the reduced model and permutation under the full model.

Unrestricted permutation of raw data defined by Manly (2007) involves unrestricted permutation of the dependent vector \mathbf{y} . In this method it is assumed that under the null hypothesis, the vector \mathbf{y} is *i.i.d* and thus exchangeable.

Permutation of residuals under the reduced model is described by Freedman and Lane (1983) and Kennedy (1995). In this method, the residuals are computed from a model that includes all parameters except the parameters of interest. The parameters of interest in this context are the independent parameters being tested for significance. This implies that only the effects of the parameters that are of interest are present in these residuals. Under the null hypothesis that the reduced model is valid, there should be no systematic effect in these residuals, and therefore an unrestricted permutation of these residuals is used to test the parameters of interest.

Permutation under the full model, that is, includes all independent variables, is described by Ter Braak (1992) and Tantawanich (2006). As the name suggests, one should first compute the residuals based on the full model, which removes the effects of all factors from the original observations. Then one constructs new samples by adding the fitted values to unrestricted permuted versions of the residuals. Based on these new samples, one uses the F – statistic whose null hypothesis is that the true parameters are equal to their empirical values in the original sample.

There is general agreement concerning an appropriate method of permutation for exact tests of hypotheses in simple linear regression, for example, methods proposed by Edgington (1995) and Manly (2007). However, this is not the case for exact partial tests in multiple linear regression. Since multiple linear regression is more complex, only recently have advancements in this area been made. An exact technique for the permutation of residuals under the reduced model has been proposed by Huh and Jung (2001) and Kherad-Pajouh and Renaud (2010). This exact method is based on the permutation of residuals under the reduced form. In this method, the residuals are computed from a model that includes all parameters, except the parameters of interest. It is based on the ideas provided by Kennedy (1995). Using a transformation on the parameters of interest, an exact test can be done using exchangeable errors.

The different permutation strategies are compared in section 9.7 to determine how well these methods perform with normal and non-normal errors. Simulations are constructed with errors that have a $N(0,1)$, $U(-\sqrt{3},\sqrt{3})$, $\exp(1)-1$ and $t(4)$ distribution. By comparing the partial F -test to permutation approach of Manly (2007), Freedman and Lane (1983), Kennedy (1995), Ter Braak (1992), Tantawanich (2006) and Kherad-Pajouh and Renaud (2010), simulations show that the partial F -test outperforms the permutation tests when errors are normal. The Kherad-Pajouh and Renaud (2010) test outperforms all the other tests when errors are non-normal because this test is an exact test and not an approximate test.

9.2 *The Parametric Approach*

Consider a multiple linear regression model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}$$

$$\mathbf{y} = \begin{pmatrix} y_1 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} x_{1,1} & \cdots & x_{1,p+q} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ x_{n,1} & \cdots & x_{n,p+q} \end{pmatrix}, \quad \beta_1 = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \cdot \\ \cdot \\ \beta_p \end{pmatrix}, \quad \beta_2 = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \cdot \\ \cdot \\ \beta_q \end{pmatrix}$$

$$\mathbf{X}_1 = \begin{pmatrix} x_{1,1} & \cdots & x_{1,p} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ x_{n,1} & \cdots & x_{n,p} \end{pmatrix}, \quad \mathbf{X}_2 = \begin{pmatrix} x_{1,1} & \cdots & x_{1,q} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ x_{n,1} & \cdots & x_{n,q} \end{pmatrix}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_n \end{pmatrix}$$

\mathbf{y} is an $n \times 1$ vector of responses, $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of *i.i.d* errors, \mathbf{X} is an $n \times (p+q)$ matrix of full rank and $\boldsymbol{\beta}$ is the vector of parameters. \mathbf{X}_1 and \mathbf{X}_2 are matrices of dimension $n \times (p+q)$ and $n \times (p+q)$ respectively and β_1 and β_2 are vectors of dimension $p \times 1$ and $q \times 1$ respectively. For any test of interest, the design matrix \mathbf{X} and the vector of parameters $\boldsymbol{\beta}$ can be divided into the component of interest, that is β_2 and the component not of interest, that is β_1 , where,

$$\mathbf{X} = [\mathbf{X}_1 \quad \mathbf{X}_2], \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}.$$

The corresponding hypotheses are

$$H_0 : \beta_2 = \mathbf{0}, \quad \text{versus } H_1 : \beta_2 \neq \mathbf{0},$$

where $\mathbf{0}$ is a $q \times 1$ vector of zeros. If the null hypothesis is not rejected, then the explanatory variables(s) in \mathbf{X}_2 may not be necessary and the reduced model in this case would be

$$\mathbf{y} = \mathbf{X}_1 \beta_1 + \boldsymbol{\varepsilon}.$$

The parametric partial F -statistic for testing $H_0 : \beta_2 = \mathbf{0}$ is of the form

$$F = \frac{\mathbf{y}' \left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1} \mathbf{X}_1' \right] \mathbf{y} / (q)}{\mathbf{y}' \left[\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right] \mathbf{y} / (n - p - q)} \quad (9.2.1)$$

Under the normality assumption, the partial F -statistic in (9.2.1) is distributed as $F_{\alpha, (q, n-p-q)}$ when $H_0 : \beta_2 = \mathbf{0}$ is true, and the null hypothesis is rejected if $F \geq F_{\alpha, (q, n-p-q)}$.

9.3 Permutation of Raw Data

9.3.1 The Manly Method

Looking at the permutation equivalent to the normal partial F -test, Manly (2007) proposes a procedure that involves the permutation of raw data. This procedure is as follows: The variable \mathbf{y} is regressed on \mathbf{X}_1 and \mathbf{X}_2 together (using least squares) to obtain an estimate $\hat{\beta}_2$ of β_2 and a value of the usual F -statistic given by equation 9.2.1. The \mathbf{y} values are then randomly permuted to obtain the permuted values \mathbf{y}^* . These \mathbf{y}^* values are then regressed on \mathbf{X}_1 and \mathbf{X}_2 (unpermuted) together to obtain an estimate $\hat{\beta}_2^*$ of β_2 and a value F^* for the permuted data. Under the null hypothesis $H_0 : \beta_2 = \mathbf{0}$, the random vector \mathbf{y} is exchangeable and thus all possible permutations of \mathbf{y} have the same distribution. The permutation statistic for Manly (2007) denoted F^{M*} is

$$F^{M*} = \frac{\mathbf{y}^{M*}' \left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1} \mathbf{X}_1' \right] \mathbf{y}^{M*} / (q)}{\mathbf{y}^{M*}' \left[\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right] \mathbf{y}^{M*} / (n - p - q)}.$$

This procedure is repeated a large number of times so that the empirical distribution of $H_0 : \beta_2 = \mathbf{0}$ can be assessed by

$$\text{p-value} = P(F^{M^*} \geq F) = \frac{\text{number of } F^{M^*} \geq F}{n!}$$

Let this test be denoted by RT1. Kennedy and Cade (1996) suggest that the permutation of raw data, as described by Manly (2007), should be controlled. The reason for this control is because tests done by Kennedy and Cade (1996) show significantly different results between permutation tests and model-based methods. This discrepancy appears when X_1 is not zero and X_1 contains an outlier. This method cannot handle these particular situations because the relationship between X_1 and Y is held constant throughout the permutations. When there are such extreme outliers in the predictor variables, these should be identifiable as high leverage points in diagnostic analyses prior to the regression analysis.

Anderson and Legendre (1999) expand on the results by Kennedy and Cades (1996) and show that the permutation of raw data results in inflated p-values when there is an extreme outlier. This is regardless of whether or not there is collinearity between predictor variables, nor if the data is normal or non-normal. This problem of inflated p-values cannot be amended by increasing the sample size as these results remain unchanged.

Outliers should therefore be removed from the data set, so that the potential problem may be eliminated beforehand. However, the presence of outliers in a multiple regression may not always be readily apparent or easy to define. Permutation methods are appreciated for their lack of assumptions concerning distributions of variables, thus diagnostic checking of distributions of variables should be unnecessary.

The method proposed by Manly (2007) is not an exact test for a partial regression coefficient in a linear model, unless all other parameters in the model are truly equal to zero. This method only provides an approximate test. Manly (2007) suggests that a test of 5000 simulations is sufficient, yet the results of Anderson and Legendre (1999) differ significantly from those of Manly (2007). There are several reasons for this discrepancy but one reason may be because Manly (2007) uses 5000 simulations, whereas Anderson and Legendre (1999) use 10,000.

9.4 Permutation under the Reduced Model

9.4.1 The Freedman and Lane Method

In contrast to the method of permuting raw data, there are those techniques which use the residuals of a linear model as the permutable units for a test. The error ε is dissociated from each value of Y by the application of a model to produce residuals, as opposed to the original Y values. The rationale for the permutation of residuals for the hypothesis $H_0: \beta_2 = \mathbf{0}$, is the following: Given some estimate of the relationship between Y and X_1 (even if it is zero) there is no further variation in Y which can be explained by X_2 . Looking at the approach proposed by Freedman and Lane (1983), the vector y is regressed on X_1 and X_2 together to obtain an estimate $\hat{\beta}_2$ of β_2 and a reference value F for the real data. To obtain the empirical distribution, the vector y is regressed on X_1 alone according to the model $y = \mu + X_1\beta_1 + \varepsilon$ providing estimates $\hat{\mu}$ of μ , $\hat{\beta}_1$ of β_1 and residuals

$$\varepsilon^{FL} = y - (\hat{\mu} + X_1\hat{\beta}_1)$$

The residuals from the regression are then permuted randomly, producing ε^{FL*} . New values for y^{FL*} are calculated by adding the permuted residuals to the fitted values as follows:

$$y^{FL*} = \hat{\mu} + X_1\hat{\beta}_1 + \varepsilon^{FL*}$$

y^{FL*} is then regressed on X_1 and X_2 together, according to the model $E(y^{FL*}) = \mu^* + X_1\beta_1^* + X_2\beta_2^*$ to obtain an estimate $\hat{\beta}_2^*$ of β_2^* , and a value F^{FL*} . Thus the permutation statistic for Freedman and Lane (1983) is given by

$$F^{FL*} = \frac{y^{FL*} \left[X(X'X)^{-1}X' - X_1(X_1'X_1)^{-1}X_1' \right] y^{FL*} / (q)}{y^{FL*} \left[I_n - X(X'X)^{-1}X' \right] y^{FL*} / (n - p - q)}$$

This procedure is repeated a large number of times so that the empirical distribution of $H_0 : \beta_2 = \mathbf{0}$ can be assessed by

$$\text{p-value} = P(F^{FL*} \geq F) = \frac{\text{number of } F^{FL*} \geq F}{n!}$$

Let this test be denoted by RT2. Permutation of the residuals under the model preserves the covariances between \mathbf{y} and X_1 , X_2 and X_1 , and among the X_1 variables, but not between \mathbf{y} and X_2 , across all permutations.

Freedman and Lane (1983) claim their method is a ðnon-stochasticö approach, referring to the proportion of the values $F^{FL*} \geq F$ as a ðdescriptive statisticö instead of a probability. However, according to Kennedy (1995) their rationale for the test is effectively that of a model-based approach. The goal is to isolate the test of \mathbf{Y} on X_2 alone, while taking X_1 into account through the use of the linear regression equation and permutation of residuals.

Freedman and Lane (1983) emphasize three conditions for the use of their method. The one condition is that the data should not contain extreme outliers and the second is that X_1 and X_2 should not be highly collinear. The third is that the sample size n should be relatively large. Since the permutation is on the residuals, the test is not an exact test in a randomization sense, but has asymptotically exact significance levels.

9.4.2 The Kennedy Method

Kennedy (1995) presented a method of permutation which he states is identical to the Freedman and Lane (1983) procedure. The rationalization for the method of Kennedy (1995) is the same as that of Freedman and Lane (1983) approach, but differs computationally. The method is described as follows: The variable \mathbf{y} is regressed on X_1 and X_2 together to obtain an estimate $\hat{\beta}_2$ of β_2 and a reference value F for the real data. To obtain the empirical distribution, the variable \mathbf{y} is regressed on X_1 and X_2 according to the model

$\mathbf{y} = \mu + \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2 + \varepsilon$. Let $\mathbf{H}_1 = \mathbf{X}_1(\mathbf{X}_1' \mathbf{X}_1)^{-1} \mathbf{X}_1'$ and multiplying $\mathbf{I}_n - \mathbf{H}_1$ to both sides of this model, one gets

$$\mathbf{y}^K = \mu + \mathbf{X}_2^K \beta_2 + \varepsilon^K$$

where

$$\mathbf{y}^K = (\mathbf{I}_n - \mathbf{H}_1)\mathbf{y}, \quad \mathbf{X}_2^K = (\mathbf{I}_n - \mathbf{H}_1)\mathbf{X}_2, \quad \varepsilon^K = (\mathbf{I}_n - \mathbf{H}_1)\varepsilon.$$

These residuals ε^K are then permuted randomly, producing ε^{K*} . New values for \mathbf{y}^{K*} are calculated by adding the permuted residuals to the fitted values as follows:

$$\mathbf{y}^{K*} = \hat{\mu} + \mathbf{X}_2^K \hat{\beta}_2 + \varepsilon^{K*}$$

\mathbf{y}^{K*} is then regressed on \mathbf{X}_2 , according to the model $E(\mathbf{y}^{K*}) = \mu + \mathbf{X}_2^K \beta_2^*$ to obtain a value F^{K*} . Thus the permutation statistic for Kennedy (1995) is given by

$$F^{K*} = \frac{\mathbf{y}^{K*}' [\mathbf{X}(\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' - \mathbf{X}_1(\mathbf{X}_1' \mathbf{X}_1)^{-1} \mathbf{X}_1'] \mathbf{y}^{K*} / (q)}{\mathbf{y}^{K*}' [\mathbf{I}_n - \mathbf{X}(\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}'] \mathbf{y}^{K*} / (n - p - q)}$$

This procedure is repeated a large number of times so that the empirical distribution of $H_0 : \beta_2 = \mathbf{0}$ can be assessed by

$$\text{p-value} = P(F^{K*} \geq F) = \frac{\text{number of } F^{K*} \geq F}{n!}$$

Let this test be denoted by RT3. According to simulations done by Anderson and Legendre (1999), the Kennedy (1995) method will not give the same results as those by Freedman and Lane (1983). The estimate of the slope coefficient vector $\hat{\beta}_2^*$ is the same, but the value of the F -statistic under permutation is different for the two methods. The reason for this dissimilarity between the two methods is subtle but has important consequences. The

Kennedy (1995) method removes the effect of the variable which is not of interest. Thus the parameters associated with X_1 remain fixed throughout the permutation procedure. In the Freedman and Lane (1983) method, this parameter does not stay fixed. The permuted residuals ε^{FL*} are added back onto the fitted values to obtain Y^* . These are then regressed on X_1 and X_2 together, so the parameter estimates for X_1 in the multiple regression model changes with each permutation. If the true values for β_1 were known, there would be no difference between the two methods. Although there is no relationship between ε^{FL} and X_1 , some small relationship is reintroduced between ε^{FL} and X_1 , by the permutation of these residuals. The method of Freedman and Lane (1983) takes this into account by maintaining the conditioning on X_1 throughout the permutation procedure, whereas that of Kennedy (1995) does not.

Kennedy (1995) states that the method of Manly (2007), that is, permuting y , is only justified when the covariable's parameter β_1 is zero. The argument is essentially that the permutation of raw data ignores the covariable parameters, which often may not be justified. Kennedy (1995) suggests that the method of permuting raw data for the test of the hypothesis $H_0 : \beta_2 = 0$ will give biased results if the errors ε and the y values have radically different distributions in the presence of a non-zero β_1 . In limited simulations, Kennedy (1995) found that the permutation of the raw data y in multiple regression resulted in an inflated p-value when outliers were included in X_1 and $\beta_1 \neq 0$. The results of Kennedy (1995) is not supported by further simulations published by Manly (2007), although Manly (2007) suggests that a more extensive simulations are needed on this topic, his results seems to show that the method of permuting y for tests of partial regression coefficients is not necessarily flawed in the way that Kennedy (1995) claims.

Simulations done by Anderson and Legendre (1999) show that the Kennedy (1995) procedure has inflated p-values, especially with small sample sizes. Permutation under the reduced model should therefore be done using the Freedman and Lane (1983) method when n is small.

9.5 Permutation under the Full Model

9.5.1 The Ter Braak Method

The two methods that have been proposed by Freedman and Lane (1983) and Kennedy (1995) have been called permutation under the ðnull modelö by Ter Braak (1992) or ðunder the reduced modelö by Cade and Richards (1996). Permutation of residuals under the full model, that is, all independent variables, was developed by Ter Braak (1992). It was introduced as the permutational analog (resampling without replacement) to the bootstrapping method (resampling with replacement). Ter Braak (1992) refers to his method as permutation ðunder the full modelö and this method uses the residuals from the full regression model as the permutable units for the test. The estimate F^* as well as the original estimate of $\hat{\beta}_2$ are used as part of the permutation procedure. Ter Braak (1992) claims that this procedure should have the effect of reducing the variance of the parameter of interest under permutation, thus increasing the power of the test. This procedure is described as follows: The vector \mathbf{y} is regressed on X_1 and X_2 together to obtain estimates $\hat{\mu}$ of μ , $\hat{\beta}_1$ of β_1 , $\hat{\beta}_2$ of β_2 and residuals ε , as well as the reference value F for the original data. The residuals ε^{TB} are calculated from the equation

$$\varepsilon^{TB} = \mathbf{y} - (\hat{\mu} + X_1\hat{\beta}_1 + X_2\hat{\beta}_2),$$

which are permuted randomly, producing ε^{TB*} . The new \mathbf{y}^{TB*} values are calculated from the permuted residuals as follows:

$$\mathbf{y}^{TB*} = \hat{\mu} + X_1\hat{\beta}_1 + X_2\hat{\beta}_2 + \varepsilon^{TB*}$$

The new values \mathbf{y}^{TB*} are regressed on X_1 and X_2 to obtain an estimate $\hat{\beta}_2$ and a value F^{TB*} under permutation. The permutation statistic for Ter Braak (1992) denoted F^{TB} is given by

$$F^{TB*} = \frac{\mathbf{y}^{TB*} \left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - X_1(X_1'X_1)^{-1}X_1' \right] \mathbf{y}^{TB*} / (q)}{\mathbf{y}^{TB*} \left[\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \right] \mathbf{y}^{TB*} / (n-p-q)}$$

This procedure is repeated a large number of times so that the empirical distribution of $H_0 : \beta_2 = \mathbf{0}$ can be assessed by

$$\text{p-value} = P(F^{TB*} \geq F) = \frac{\text{number of } F^{TB*} \geq F}{n!}$$

Let this test is denoted by RT4. Permutation of the residuals under the null model preserves all covariances across the permutations, that is, among y , X_1 and X_2 , as well as the X_1 variables. According to Ter Braak (1992), this permutation test under the full model has asymptotically exact significance levels. A point to note about this approach is that the F – statistic is calculated under permutation according to the hypothesis that $\hat{\beta}_2^{**} = \hat{\beta}_2^{**}$, that is, that the values of $\hat{\beta}_2^{**}$ obtained under permutation are close to the original estimated values of $\hat{\beta}_2^{**}$. For this reason, this approach has also been called permutation under the alternative hypothesis by Ter Braak (1992).

According to Anderson and Legendre (1999), permutation under the reduced model of Freedman and Lane (1983) and the full model of Ter Braak (1992) generally give similar results and are equally appropriate for most situations, with specific reference to univariate models. In the extreme situation of a remote outlier in the variable X_1 with extremely non-normal errors and small sample sizes, the Ter Braak (1992) method may be destabilized. But this instability disappears when n is large, ($n=100$) or more reasonable error structures are used, such as the normal or exponential. The introduction of an outlier in X_1 has no effect on the level accuracy of the Freedman and Lane (1983) method of permutation, in any situation.

Although the Freedman and Lane (1983) method might be preferable to use with smaller sample sizes, there is a computational advantage in using the Ter Braak (1992) method. One can use the permutation of a single set of residuals from the full model to test a number of different hypotheses concerning individual partial regression coefficients in a multiple regression model. The Freedman and Lane (1983) method is computationally more intensive. When testing several different hypotheses about different coefficients in multiple regression,

the Freedman and Lane (1983) method requires different sets of residuals from several reduced models.

9.5.2 The Tantanich Method

Tantanich (2006) proposes an alternative permutation procedure which is also under the full model. This procedure is similar to the Ter Braak (1992) method, but instead of calculating new y values from the permuted residuals, new β_2 values are calculated. The estimate F^* as well as the original estimate of β_2 are used as part of the permutation procedure. Tantanich (2006) claims that this method provides a higher power than the methods proposed by Ter Braak (1992) and Manly (2007). The procedure is described as follows: Let $H = X(X'X)^{-1}X'$ where H is a $n \times n$ matrix and let $H_1 = X_1(X_1'X_1)^{-1}X_1'$, where H_1 is a $n \times n$ matrix. For the OLS estimator of β_2 ,

$$\begin{aligned}
 \hat{\beta}_2 &= [I_n - X_1(X_1'X_1)^{-1}X_1']X_1 \\
 &= I_n X_1 - X_1(X_1'X_1)^{-1}X_1'X_1 \\
 &= X_1 - X_1(X_1'X_1)^{-1}(X_1'X_1) \\
 &= X_1 - X_1 I_p \\
 &= X_1 - X_1 \\
 &= \mathbf{0}
 \end{aligned}$$

One can write

$$\hat{\beta}_2 - \beta_2 = X_2(X_2'X_2)^{-1}X_2' - \beta_2$$

$$\begin{aligned}
&= [\mathbf{X}_2'(\mathbf{I}_n - \mathbf{H}_1)\mathbf{X}_2]^{-1} \mathbf{X}_2'(\mathbf{I}_n - \mathbf{H}_1)\mathbf{y} - \beta_2 \\
&= [\mathbf{X}_2'(\mathbf{I}_n - \mathbf{H}_1)\mathbf{X}_2]^{-1} \mathbf{X}_2'(\mathbf{I}_n - \mathbf{H}_1)\mathbf{y} - \mathbf{0} - \mathbf{I}_q \beta_2 \\
&= [\mathbf{X}_2'(\mathbf{I}_n - \mathbf{H}_1)\mathbf{X}_2]^{-1} \mathbf{X}_2'(\mathbf{I}_n - \mathbf{H}_1)\mathbf{y} - [\mathbf{X}_2'(\mathbf{I}_n - \mathbf{H}_1)\mathbf{X}_2]^{-1} \mathbf{X}_2'(\mathbf{I}_n - \mathbf{H}_1)\mathbf{X}_1 \beta_1 \\
&\quad - [\mathbf{X}_2'(\mathbf{I}_n - \mathbf{H}_1)\mathbf{X}_2]^{-1} \mathbf{X}_2'(\mathbf{I}_n - \mathbf{H}_1)\mathbf{X}_2 \beta_2 \\
&= [\mathbf{X}_2'(\mathbf{I}_n - \mathbf{H}_1)\mathbf{X}_2]^{-1} \mathbf{X}_2'(\mathbf{I}_n - \mathbf{H}_1)[\mathbf{y} - (\mathbf{X}_1 \beta_1 + \mathbf{X}_2 \beta_2)] \\
&= [\mathbf{X}_2'(\mathbf{I}_n - \mathbf{H}_1)\mathbf{X}_2]^{-1} \mathbf{X}_2'(\mathbf{I}_n - \mathbf{H}_1)\boldsymbol{\varepsilon}
\end{aligned} \tag{9.5.1}$$

The residuals $\boldsymbol{\varepsilon}^T$ are calculated from

$$\boldsymbol{\varepsilon}^T = \mathbf{y} - (\mathbf{X}_1 \hat{\beta}_1 + \mathbf{X}_2 \hat{\beta}_2),$$

and used as the estimator of the vector of errors and randomly permuted, producing $\boldsymbol{\varepsilon}^{T*}$. A new permuted estimator for β_2 is calculated from the permuted residuals and equation (9.5.1), thus

$$\hat{\beta}_2^{T*} = \hat{\beta}_2 + [\mathbf{X}_2'(\mathbf{I}_n - \mathbf{H}_1)\mathbf{X}_2]^{-1} \mathbf{X}_2'(\mathbf{I}_n - \mathbf{H}_1)\boldsymbol{\varepsilon}^{T*}$$

Multiplying the partial F -statistic by $q/n - p - q$ yields the same result as omitting the degrees of freedom and is easier to calculate. Re-writing the F -statistic in terms of $\hat{\beta}_2$, $(\mathbf{I}_n - \mathbf{H}_1)$ and \mathbf{H} ,

$$F^T = \frac{\hat{\beta}_2' [\mathbf{X}_2'(\mathbf{I}_n - \mathbf{H}_1)\mathbf{X}_2] \hat{\beta}_2}{\mathbf{y}' \mathbf{H} \mathbf{y}} \tag{9.5.2}$$

Substituting $\hat{\beta}_2$ in equation 9.5.2 by $\hat{\beta}_2^{T*} = \hat{\beta}_2 + [\beta_1'(\mathbf{I}_n - \mathbf{H}_1)\beta_1]^{-1} \beta_1'(\mathbf{I}_n - \mathbf{H}_1)\boldsymbol{\varepsilon}^*$, the statistic F^T becomes a permuted statistic F^{T*} , which is defined as

$$F^{T*} = \frac{\hat{\beta}_2^{T*} [\mathbf{X}_2'(\mathbf{I}_n - \mathbf{H}_1)\mathbf{X}_2] \hat{\beta}_2^{T*}}{\mathbf{y}'(\mathbf{I}_n - \mathbf{H}_1)\mathbf{y}}$$

Thus the proposed sampling permutation test of $H_0 : \beta_2 = \mathbf{0}$, an (estimated) empirical distribution of F^T is obtained from $n!$ permutation statistics. Thus the empirical distribution of $H_0 : \beta_2 = \mathbf{0}$ can be assessed by

$$\text{p-value} = P(F^{T*} \geq F) = \frac{\text{number of } F^{T*} \geq F}{n!} \leq \alpha$$

Let this test is denoted by RT5. As with Ter Braak (1992), the F -statistic is also calculated under permutation according to the hypothesis that $\beta_2^{**} = \beta_2^{**}$. Simulations done by Tantawanich (2006) show that the proposed method performs as well as the partial F -test for the normal error case when the sample size is large. It outperforms the methods proposed by Manly (2007) and Ter Braak (1992) when the sample size is small. For the non-normal error case, the proposed test has the highest power for small sample sizes. However, as the sample size increases, there is little difference between the Tantawanich (2006), Manly (2007) and Ter Braak (1992) methods.

9.6 Permutation for an Exact Test

9.6.1 The Kherad-Pajouh and Renaud Method

All the methods described in sections 9.3, 9.4 and 9.5 are approximate permutation tests because there is some correlation between the residuals. To achieve an exact permutation test, this correlation should be removed to ensure that the errors are exchangeable. Kherad-Pajouh and Renaud (2010) provide a procedure that is an exact permutation test. Using the residuals under the reduced model, the idea is to remove that part of the design matrix that is not tested. The method is based on the article written by Jung et al. (2001). In the article, Jung et al (2001) proposes an exact permutation procedure for only the highest-order factor. The hypothesis $H_0 : \beta_p = 0$ for $\beta_1, \beta_2, \dots, \beta_{p-1}, \beta_p$ is tested. Kherad-Pajouh and Renaud (2010) extend the Jung et al. (2001) method to a more general approach for partial multivariate regression.

An advantage of this exact method compared to methods using restriction of raw data is that this method is also applicable to designs with small sample sizes. This is because the

proposed method relies on a large number of permutations. Kherad-Pajouh and Renaud (2010) use the residuals under the modified model approach and show that they satisfy the exchangeability condition, implying that they lead to an exact permutation test.

Exchangeability under the null hypothesis is the only requirement for having an exact permutation test. Thus Kherad-Pajouh and Renaud (2010) provide an exact permutation strategy by modifying the approximate test proposed by Kennedy (1995). The main idea is to remove the correlation between residuals, using the decomposition of an idempotent matrix, thus obtaining *i.i.d* or exchangeable errors. Kherad-Pajouh and Renaud (2010) show that, if the error terms are *i.i.d* or exchangeable, the modified residuals are exchangeable up to the second moment.

The condition to obtain a permutation test is to handle exchangeable objects. The elements in \mathbf{y} for $\mathbf{y} = \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2 + \varepsilon$ are not exchangeable. Since under the null hypothesis of $H_0 : \beta_2 = \mathbf{0}$, $E(\mathbf{y}) = \mathbf{X}_1\beta_1$ differs for each \mathbf{y} . Using the residuals under the reduced model solves this problem. Part of the design matrix which is not of interest is removed for the test. Let $\mathbf{H}_1 = \mathbf{X}_1(\mathbf{X}_1' \mathbf{X}_1)^{-1} \mathbf{X}_1'$ where \mathbf{H}_1 is an $n \times n$ matrix. Multiplying \mathbf{H}_1 to both sides of the equation $\mathbf{y} = \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2 + \varepsilon$, one gets

$$\begin{aligned} [\mathbf{I}_n - \mathbf{H}_1]\mathbf{y} &= [\mathbf{I}_n - \mathbf{H}_1](\mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2 + \varepsilon) \\ [\mathbf{I}_n - \mathbf{H}_1]\mathbf{y} &= [\mathbf{I}_n - \mathbf{X}_1(\mathbf{X}_1' \mathbf{X}_1)^{-1} \mathbf{X}_1'](\mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2) + [\mathbf{I}_n - \mathbf{H}_1]\varepsilon \\ [\mathbf{I}_n - \mathbf{H}_1]\mathbf{y} &= \mathbf{I}_n\mathbf{X}_1\beta_1 - \mathbf{X}_1(\mathbf{X}_1' \mathbf{X}_1)^{-1} \mathbf{X}_1' \mathbf{X}_1\beta_1 + [\mathbf{I}_n - \mathbf{X}_1(\mathbf{X}_1' \mathbf{X}_1)^{-1} \mathbf{X}_1']\mathbf{X}_2\beta_2 + [\mathbf{I}_n - \mathbf{H}_1]\varepsilon \\ [\mathbf{I}_n - \mathbf{H}_1]\mathbf{y} &= \mathbf{X}_1\beta_1 - \mathbf{X}_1\mathbf{I}_q\beta_1 + [\mathbf{I}_n - \mathbf{H}_1]\mathbf{X}_2\beta_2 \\ [\mathbf{I}_n - \mathbf{H}_1]\mathbf{y} &= [\mathbf{I}_n - \mathbf{H}_1]\mathbf{X}_2\beta_2 + [\mathbf{I}_n - \mathbf{H}_1]\varepsilon \\ \mathbf{y}^{KR1} &= \mathbf{X}_2^{KR1}\beta_2 + \varepsilon^{KR1} \end{aligned}$$

where

$$\mathbf{y}^{KR1} = [\mathbf{I}_n - \mathbf{H}_1]\mathbf{y}, \quad \mathbf{X}_2^{KR1} = [\mathbf{I}_n - \mathbf{H}_1]\mathbf{X}_2, \quad \varepsilon^{KR1} = [\mathbf{I}_n - \mathbf{H}_1]\varepsilon,$$

In a balanced case, since $X_1'X_2 = \mathbf{0}$, it can be seen that $X^{KR1} = X_2$. Then the least squares estimate is given by

$$\beta_2^{KR1} = (X_2^{KR1} X_2^{KR1})^{-1} X_2^{KR1} y^{KR1}$$

It can be seen that β_2^{KR1} is equal to the least squares estimate $\hat{\beta}_2$. As in the Kennedy (1995) method, these residuals ε^{KR1} are permuted randomly, producing ε^{KR1*} . New values for y^{KR1*} are calculated by adding the permuted residuals to the fitted values as follows:

$$y^{KR1*} = \hat{\mu} + X_2^{KR1} \hat{\beta}_2 + \varepsilon^{KR1*}$$

y^{KR1*} is then regressed on X_2 to obtain the permutation statistic F^{KR1*} expressed as

$$F^{KR1*} = \frac{y^{KR1*} \left[X^{KR1} (X^{KR1} X^{KR1})^{-1} X^{KR1} \right] y^{KR1*} / (q)}{y^{KR1*} \left[I_n - X(X'X)^{-1} X' \right] y^{KR1*} / (n-p-q)} \quad (9.6.1)$$

The proof is given in the appendix. This procedure is not an exact permutation approach because $\varepsilon^{KR1} \sim (\mathbf{0}, \sigma^2 (I_n - H_1))$. For an exact method, these errors should be exchangeable. Exchangeability concerns the empirical distribution that is obtained from the permutations, as well as the variances, covariances and higher order moments. Since $\varepsilon^{KR1} \sim (\mathbf{0}, \sigma^2 (I_n - H_1))$, under the null hypothesis, $H_0 : \beta_2 = \mathbf{0}$, the distribution of the randomly permuted vector y^{KR1*} is distributed differently from y^{KR1} .

Since the matrix X_1 has full rank, there exists a matrix V of dimension $n \times q$ whose columns form an orthonormal basis for the subspace orthogonal to $\text{span}(X_1)$. Orthonormality of the columns of V implies that $V'V = I_q$ where

$$\begin{pmatrix} v_{1,1} & \cdots & v_{1,n} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ v_{q,1} & \cdots & v_{q,n} \end{pmatrix} \begin{pmatrix} v_{1,1} & \cdots & v_{1,q} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ v_{n,1} & \cdots & v_{n,q} \end{pmatrix} = \begin{pmatrix} 1 & \cdots & 0 \\ \cdot & 1 & \cdot \\ \cdot & & \cdot \\ \cdot & & 1 \cdot \\ 0 & \cdots & 1 \end{pmatrix}$$

since $v_{ik} = \begin{cases} 1 & \text{for } k=i \\ 0 & \text{for } k \neq i \end{cases}$ The subspaces spanned by X_1 and V are complementary by

construction, thus the sum of the corresponding projections are equal to the identity matrix.

Since

$$I_n = X_1(X_1'X_1)^{-1}X_1' + V(V'V)^{-1}V'$$

one gets,

$$I_n = X_1(X_1'X_1)^{-1}X_1' + V(I_q)^{-1}V'$$

$$I_n = X_1(X_1'X_1)^{-1}X_1' + VV'$$

$$I_n = H_1 + VV'$$

$$VV' = I_n - H_1$$

Thus the matrix V satisfies the following equations:

$$VV' = I_n - H_1 \text{ and } V'V = I_p \quad (9.6.2)$$

The construction of the matrix V depends on the decomposition of $(I_n - H_1)$ into eigenvalues and eigenvectors. Since $(I_n - H_1)$ is a symmetric and idempotent matrix, it possesses only two distinct eigenvalues, zero and one. Let $(I_n - H_1) = UDU$ be the eigenvalue and eigenvector decomposition of $(I_n - H_1)$, where D is the diagonal matrix containing the eigenvalues of $(I_n - H_1)$ and U the unitary matrix, whose columns are the eigenvectors of $(I_n - H_1)$. Since $(I_n - H_1)$ has rank $n-p$, there are $n-p$ ones and p zeroes in the diagonal of D .

Both conditions of (9.6.2) are satisfied by choosing V as the columns of U corresponding to the non-zero diagonal elements of D . The eigenvectors corresponding to the zero eigenvalue do not contribute in the product, and therefore $VV' = UDU' = I_n - H_1$. Since U is unitary, it implies $VV' = I_n$.

There is an $n-p$ dimensional subspace which is mapped to itself under the projection induced by $(I_n - H_1)$. Any orthonormal basis of such subspace can be chosen as the eigenvectors of $(I_n - H_1)$ corresponding to the non-zero eigenvalues. However according to Jung et al. (2006), different choices of such basis vectors result in different performance of the test in terms of level and power. Therefore, it is worth considering some specifications for choices which give better performance.

By multiplying V' to both sides of the equation $\mathbf{y}^{KR1} = X_2^{KR1} \beta_2 + \varepsilon^{KR1}$, one gets

$$\begin{aligned} V' \mathbf{y}^{KR1} &= V' (X_2^{KR1} \beta_2 + \varepsilon^{KR1}) \\ V' \mathbf{y}^{KR1} &= V' X_2^{KR1} \beta_2 + V' \varepsilon^{KR1} \\ V' \mathbf{y} &= V' X_2^{KR1} \beta_2 + V' \varepsilon^{KR1} \\ \mathbf{y}^{KR2} &= X_2^{KR2} \beta_2 + \varepsilon^{KR2} \end{aligned}$$

where

$$\mathbf{y}^{KR2} = V' \mathbf{y}^{KR1}, \quad X_2^{KR2} = V' X_2^{KR1}, \quad \varepsilon^{KR2} = V' \varepsilon^{KR1}.$$

Since

$$\text{var}(\varepsilon^{KR2}) = \text{var}(V' \varepsilon^{KR1}) = V' \text{var}(\varepsilon^{KR1}) V = V' (\sigma^2 (I - H_1)) V = \sigma^2 V' V V' V = \sigma^2 I_{n-p},$$

exchangeability of the errors are thus satisfied since $\varepsilon^{KR2*} \sim (\mathbf{0}, \sigma^2 I_n)$. Under the null hypothesis, $H_0 : \beta_2 = \mathbf{0}$, the distribution of the randomly permuted vector \mathbf{y}^{KR2*} is the same as \mathbf{y}^{KR2} . Since

$$X_2^{KR2*} \mathbf{y}^{KR2*} = X_2^{KR1} V' V \mathbf{y}^{KR1} = X_2^{KR1} (I_n - H_1) \mathbf{y}^{KR1}$$

and

$$\mathbf{X}_2^{KR2} \mathbf{X}_2^{KR2} = \mathbf{X}_2^{KR1} \mathbf{V}_1' \mathbf{V}_1 \mathbf{X}_2^{KR1} = \mathbf{X}_2^{KR1} (\mathbf{I}_n - \mathbf{H}_1) \mathbf{X}_2^{KR1}$$

one gets

$$\hat{\beta}_2^{\hat{\hat{}}} = \beta_2^{KR1} = \beta_2^{KR2} = (\mathbf{X}_2^{KR2} \mathbf{X}_2^{KR2}) \mathbf{X}_2^{KR2} \mathbf{y}^{KR2}.$$

It can be seen that β_2^{KR2} is equal to the least squares estimate $\hat{\beta}_2^{\hat{\hat{}}}$. The residuals ε^{KR2} are permuted randomly, producing ε^{KR2*} and new values for \mathbf{y}^{KR2*} are calculated by adding the permuted residuals to the fitted values as follows:

$$\mathbf{y}^{KR2*} = \hat{\mu} + \mathbf{X}_2^{KR2} \beta_2 + \varepsilon^{KR2*}. \quad (9.6.3)$$

\mathbf{y}^{KR2*} is then regressed on \mathbf{X}_2 to obtain the permutation statistic F^{KR2*} expressed as

$$F^{KR2} = \frac{\mathbf{y}' \left[\mathbf{X}^{KR2} (\mathbf{X}^{KR2} \mathbf{X}^{KR2})^{-1} \mathbf{X}^{KR2} \right] \mathbf{y} / (q)}{\mathbf{y}' \left[\mathbf{I}_n - \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \right] \mathbf{y} / (n - p - q)}$$

For the observed data, the three F statistics are equivalent:

$$F = F^{KR1} = F^{KR2} \quad (9.6.4)$$

The proof is given in the appendix. Since the equalities $\hat{\beta}_2^{\hat{\hat{}}} = \beta_2^{KR1} = \beta_2^{KR2}$ are true and $F = F^{KR1} = F^{KR2}$ have been determined, the empirical distribution of $H_0 : \beta_2 = \mathbf{0}$ can be assessed by

$$\text{p-value} = P(F^{KR2*} \geq F^{KR2}) = \frac{\text{number of } F^{KR2*} \geq F^{KR2}}{n!} \leq \alpha$$

Let this test is denoted by RT6. By using the decomposition of an idempotent matrix, the data is transformed to make the residuals exchangeable up to the second moment. This exact permutation procedure can be used for balanced and unbalanced designs, for single error terms. Simulation studies done by Kherad-Pajouh and Renaud (2010) show that that this method is competitive in terms of power and thus shows good efficiency. This proposed method is compared to the approximate methods of Manly (2007) and Freedman and Lane (1986) and results show that this proposed method performs better when dealing with small sample sizes.

9.7 Simulations

Consider a model

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \varepsilon_i$$

where $i = 1, 2, \dots, n$. For each i , x_{i1} , x_{i2} , x_{i3} and x_{i4} are generated from a uniform distribution with mean 1 and variances 2, 3, 5 and 6 respectively. ε_i is generated from either a $N(0,1)$, $U(-\sqrt{3}, \sqrt{3})$, $\exp(1)-1$ or $t(4)$. Hence y_i can be calculated for each combination of $(\beta_1, \beta_2, \beta_3, \beta_4)$. In this study, $(\beta_1, \beta_2) = (1, 1)$, $(2, 2)$; $(\beta_3, \beta_4) = (0, 0)$, $(0.1, 0.1)$, $(0.1, 0.3)$, $(0.1, 0.12)$; and $n = 8, 12$ and 24 are given. This simulation is the same as that simulation performed by Tantawanich (2006) for comparing his own approach to the Manly (2007) and Ter Braak (1992) methods. The simulation being done here expands the Tantanawich (2006) approach by comparing the permutation of the raw data, permutation under the reduced model, permutation under the full model, and the recent approach, an exact permutation method. As done by Anderson and Legendre (1999), the permuted sampling number is chosen to be 999 and the number of simulation runs for each combination is 10,000 runs for each of the six different permutation methods: Manly (2007), Freedman and Lane (1983), Kennedy (1995), Ter Braak (1992), Tantanawich (2006) and Kherad-Pajouh and Renaud (2010). The probability associated with the normal-theory partial F -test is also calculated for each data set.

The empirical type I error is estimated from the rejection rates out of 10,000 when the null hypothesis is true, that is, $(\beta_3, \beta_4) = (0, 0)$. Tables 9.7.1 and 9.7.2 exhibit type I errors of the several permutation methods as well as the partial F -test. Furthermore, the values are used for testing whether the type I error is statistically different from $\alpha = 0.05$. The Kennedy (1995) method results in an inflated type I error for errors from the normal, uniform, exponential and t distribution. This is especially apparent with small sample sizes, with the problem decreasing as the sample size increases. The results from the Kennedy (1995) method are similar to the simulation results obtained by Anderson and Legendre (1999). According to the results of Anderson and Legendre (1999) the presence of non-zero parameters for β_2 , or the presence of collinearity between the independent variables has little influence on the inflated type I error at small sample sizes. The results from the Manly (2007) method show that for normal and non-normal errors, error rates of 0.05 are maintained when sample sizes are small. With non-normal errors for the Freedman and Lane (1986), Ter Braak (1992) and Tantanawich (2006), results are conservative at small sample sizes. It is only as the sample size increases that these methods converge to produce similar results. This convergence validates those simulations performed by Anderson and Legendre (1999). Their simulations comparing the Freedman and Lane (1986) and Ter Braak (1992) methods show that these methods converge asymptotically to an appropriate type I error much more quickly than the normal partial F -test.

For the normal and exponential errors, there are no significant differences among all the methods, except for the Kennedy (1995) method. They match the normal theory partial F -test and have a type I error which does not differ significantly from 0.05 in all sets of simulations. In contrast, with increases in the covariables parameter, permutation with the Freedman and Lane (1983) method has the best level accuracy for smaller sample sizes. The Kherad-Pajouh and Renaud (2010) method has the best level of accuracy for small, medium and large sample sizes. The accuracy of the Kherad-Pajouh and Renaud (2010) is expected since it is an exact permutation method rather than an approximate approach. Comparing the permutation results to the partial F -test, one can see that the permutation results outperform the All permutation methods converge to an appropriate type I error much more quickly than the normal-theory partial F -test in situations of extremely non-normal error distributions.

Table 9.7.1 Rejection rates (out of 10,000 simulations) of the six tests when $(\beta_1, \beta_2) = (0,0)$ at $\alpha = 0.05$ with normal and uniform distributions of errors and difference values of (β_1, β_2) for sample sizes $n = 8, 12$ and 24 .

| Tests | n | $\varepsilon_i \sim N(0,1)$ $(\beta_1, \beta_2) = (1,1)$ | $\varepsilon_i \sim N(0,1)$ $(\beta_1, \beta_2) = (2.4, 2.4)$ | $\varepsilon_i \sim U(-\sqrt{3}, \sqrt{3})$ $(\beta_1, \beta_2) = (1,1)$ | $\varepsilon_i \sim U(-\sqrt{3}, \sqrt{3})$ $(\beta_1, \beta_2) = (2.4, 2.4)$ |
|-----------------------------------|-----|---|--|---|--|
| Partial F -test | 8 | 0.0497 | 0.0497 | 0.0343 | 0.0291 |
| | 12 | 0.0497 | 0.0495 | 0.0437 | 0.0378 |
| | 24 | 0.0501 | 0.0502 | 0.0483 | 0.0478 |
| RT1 (Manly Test) | 8 | 0.0470 | 0.0451 | 0.0471 | 0.0468 |
| | 12 | 0.0489 | 0.0478 | 0.0484 | 0.0469 |
| | 24 | 0.0503 | 0.0494 | 0.0492 | 0.0491 |
| RT2 (Freedman and Lane) | 8 | 0.0481 | 0.0475 | 0.0469 | 0.0466 |
| | 12 | 0.0489 | 0.0479 | 0.0472 | 0.0474 |
| | 20 | 0.0496 | 0.0491 | 0.0487 | 0.0488 |
| RT3 (Kennedy) | 8 | 0.0742 | 0.0772 | 0.0635 | 0.0625 |
| | 12 | 0.0621 | 0.0597 | 0.0611 | 0.0631 |
| | 24 | 0.0539 | 0.0552 | 0.0514 | 0.0519 |
| RT4 (TerBraak) | 8 | 0.0463 | 0.0451 | 0.0462 | 0.0460 |
| | 12 | 0.0489 | 0.0479 | 0.0473 | 0.0478 |
| | 24 | 0.0503 | 0.0492 | 0.0491 | 0.0493 |
| RT5 (Tantawanich) | 8 | 0.0474 | 0.0448 | 0.0432 | 0.0421 |
| | 12 | 0.0482 | 0.0473 | 0.0445 | 0.0430 |
| | 24 | 0.0496 | 0.0489 | 0.0469 | 0.0442 |
| RT6 (Kherad-Pajouh and Renaud) | 8 | 0.0510 | 0.0513 | 0.0485 | 0.0473 |
| | 12 | 0.0506 | 0.0510 | 0.0491 | 0.0483 |
| | 24 | 0.0495 | 0.0488 | 0.0502 | 0.0491 |

Table 9.7.2 Rejection rates (out of 10,000 simulations) of the six tests when $(\beta_1, \beta_2) = (0,0)$ at $\alpha = 0.05$ with exponential and t distributions of errors and difference values of (β_1, β_2) for sample sizes $n = 8, 12$ and 24 .

| Tests | n | $\varepsilon_i \sim \exp(1) - 1$ $(\beta_1, \beta_2) = (1,1)$ | $\varepsilon_i \sim \exp(1) - 1$ $(\beta_1, \beta_2) = (2.4, 2.4)$ | $t(4)$ $(\beta_1, \beta_2) = (1,1)$ | $t(4)$ $(\beta_1, \beta_2) = (2.4, 2.4)$ |
|---------------------|-----|--|---|--|---|
| Partial F -test | 8 | 0.0139 | 0.0135 | 0.0218 | 0.0211 |
| | 12 | 0.0138 | 0.0139 | 0.0316 | 0.0275 |
| | 24 | 0.0327 | 0.0227 | 0.0417 | 0.0429 |
| RT1 (Manly Test) | 8 | 0.0589 | 0.0318 | 0.0373 | 0.0370 |
| | 12 | 0.0497 | 0.0344 | 0.0399 | 0.0401 |
| | 24 | 0.0593 | 0.0369 | 0.0467 | 0.0463 |

| | | | | | |
|-----------------------------------|----|--------|--------|--------|--------|
| RT2 (Freedman and Lane) | 8 | 0.0203 | 0.0202 | 0.0326 | 0.0323 |
| | 12 | 0.0274 | 0.0258 | 0.0369 | 0.0372 |
| | 24 | 0.0337 | 0.0309 | 0.0465 | 0.0460 |
| RT3 (Kennedy) | 8 | 0.0503 | 0.0502 | 0.0526 | 0.0531 |
| | 12 | 0.0478 | 0.0483 | 0.0479 | 0.0473 |
| | 20 | 0.0472 | 0.0479 | 0.0477 | 0.0569 |
| RT4 (TerBraak) | 8 | 0.0257 | 0.0256 | 0.0283 | 0.0282 |
| | 12 | 0.0392 | 0.0288 | 0.0364 | 0.0387 |
| | 20 | 0.0419 | 0.0379 | 0.0473 | 0.0465 |
| RT5 (Tantawanich) | 8 | 0.0386 | 0.0231 | 0.0341 | 0.0339 |
| | 12 | 0.0379 | 0.0266 | 0.0349 | 0.0352 |
| | 20 | 0.0435 | 0.0417 | 0.0436 | 0.0426 |
| RT6 (Kherad-Pajouh and Renaud) | 8 | 0.0509 | 0.0512 | 0.0481 | 0.0472 |
| | 12 | 0.0482 | 0.0509 | 0.0484 | 0.0480 |
| | 24 | 0.0490 | 0.0489 | 0.0501 | 0.0486 |

Table 9.7.3 and 9.7.4 show the empirical power of the permutation tests and the partial F -test when $\alpha = 0.05$. Table 9.7.3 shows the power comparison when errors are *i.i.d* normally distributed, while table 9.7.4 shows the power comparison when the errors have a $t(4)$ distribution. The Kennedy (1995) method is not included in tests of power since this method generally has inflated type I error rates. With all methods, there are increases in power with an increase in sample size. On the other hand, for data generated with radically non-normal errors, the normal-theory F -test is significantly less powerful than the permutation methods. None of the permutation methods differ significantly in terms of power for any of the simulations. Differences between the reduced and full-model methods are detectable when sample sizes are small, but disappear as sample sizes are increased. At low values of β_3 and β_4 , the Freedman and Lane (1983) method has slightly greater power. As β_3 and β_4 increases, the Ter Braak (1992) method becomes more powerful than the Freedman and Lane (1983) method. As power approaches 100%, the two methods converge. The size of the difference also decreases as the sample size increases. Comparing the Tantawanich (2006) method to the Kherad-Pajouh and Renaud (2010) method, the Kherad-Pajouh and Renaud (2010) method is more powerful, however as power approaches 100%, the two methods converge.

Table 9.7.3 Power comparison of the six tests at $\alpha = 0.05$ for sample sizes $n = 8, 12$ and 24 and parameter values $(\beta_1, \beta_2) = (1, 1)$ when errors are i.i.d normal.

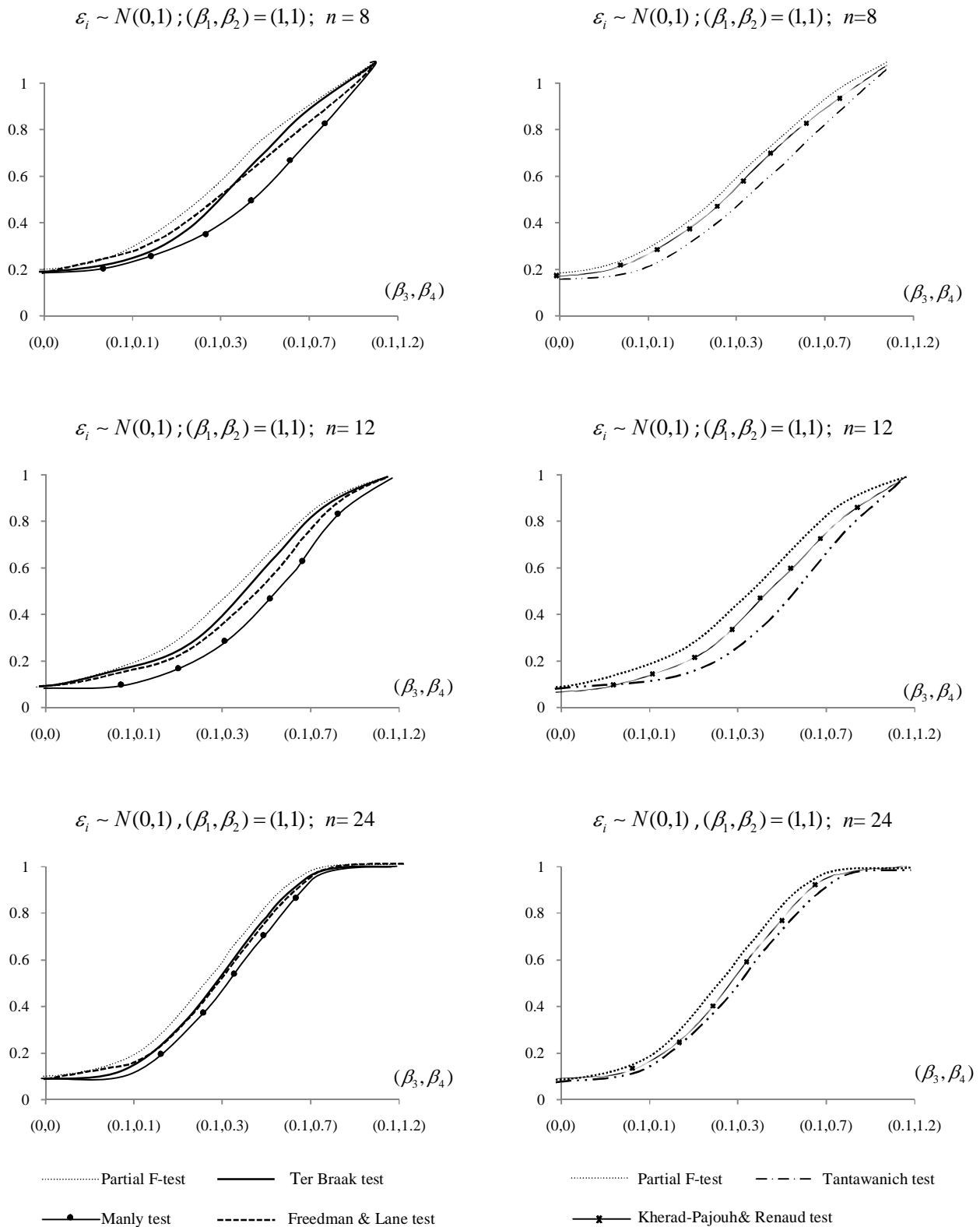
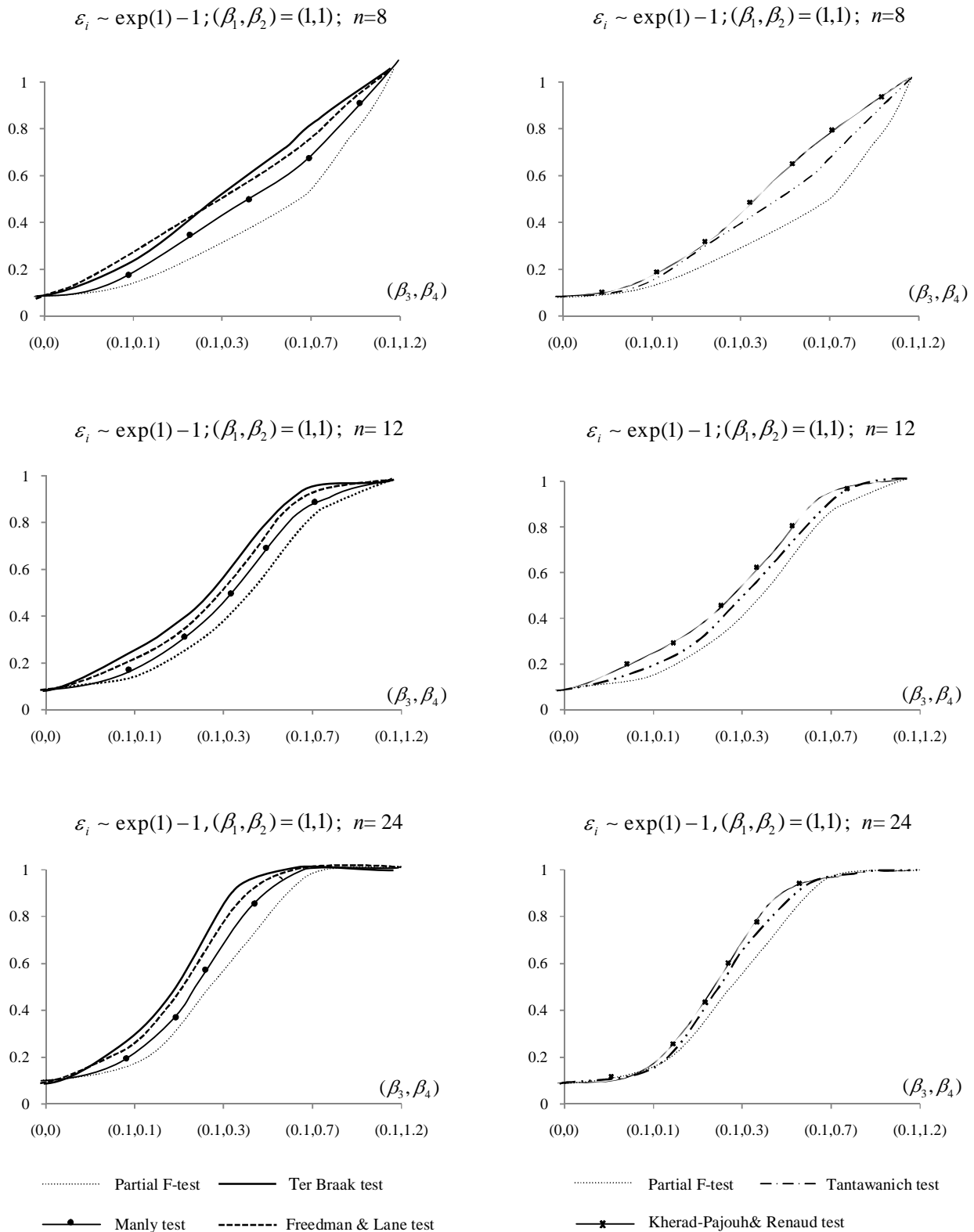


Table 9.7.4 Power comparison of the six tests at $\alpha = 0.05$ for sample sizes $n = 8, 12$ and 24 and parameter values $(\beta_1, \beta_2) = (1, 1)$ when errors are exponential.



10 Principal Component Analysis

10.1 Introduction

Principal component analysis (PCA) is an analysis method frequently used in the social sciences to reduce a large number of possibly correlated variables to a smaller number of uncorrelated underlying variables, called principal components. These orthogonal principal components contain as much information from the observed variables as possible. If the goal of the analysis is to optimally reduce a large number of variables to a smaller number, instead of deriving a model of the correlation structure, then PCA is the more appropriate procedure than compared to factor analysis. PCA can be seen as a type of exploratory analysis and its general objectives are data reduction and interpretation.

Permutation tests can be used as a means to validate and confirm the results obtained from the exploratory PCA. In PCA, the eigenvalues may be obtained using the correlation matrix. The correlation matrix requires the assumption of normality. If the data does not come from a normal distribution, the results from the PCA analysis are not valid. In PCA it is also assumed that there is a linear relationship between the variables. If this is not the case then the PCA results are also not valid. Permutation tests can thus be used to either confirm or reject the PCA results. The general rule for PCA is that most of the variation can be explained by the principal components that have eigenvalues greater than 1, otherwise known as the Kaiser Criterion. However, in real life situations, this rule is not always easy to apply because if an eigenvalue is close to 1, such as 0.988, it is debatable whether or not this eigenvalue should be dropped. Permutation tests will thus provide a definite answer for this eigenvalue.

Two permutation methods are discussed. The first method involves randomly permuting the elements in the data matrix. The second involves permuting one variable at a time, while keeping the other variables fixed. It is shown that permutation tests can always almost be performed for determining significant deviations from an alternative random explanation for the effects in the data.

10.2 The Principal Component Procedure

Algebraically principal components are particular linear combinations of the p random variables X_1, X_2, \dots, X_p . Geometrically, these linear combinations represent the selection of a new coordinate system obtained by rotating the original system with X_1, X_2, \dots, X_p as the coordinate axes. The new axes represent the directions with maximum variability and provide a simpler and more parsimonious description of the covariance structure. Principal components depend solely on the covariance matrix Σ or the correlation matrix ρ of X_1, X_2, \dots, X_p .

Let the random vector $\mathbf{X}' = [X_1, X_2, \dots, X_p]$ have the covariance matrix Σ with eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$. Consider the linear combinations

$$\begin{aligned} Y_1 &= \mathbf{a}_1' \mathbf{X} = a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p \\ Y_2 &= \mathbf{a}_2' \mathbf{X} = a_{21}X_1 + a_{22}X_2 + \dots + a_{2p}X_p \\ &\cdot \\ &\cdot \\ &\cdot \\ Y_p &= \mathbf{a}_p' \mathbf{X} = a_{p1}X_1 + a_{p2}X_2 + \dots + a_{pp}X_p \end{aligned}$$

The principal components are the uncorrelated linear combinations Y_1, Y_2, \dots, Y_p whose variances are as large as possible. The first principal component is the linear combination $\mathbf{a}_1' \mathbf{X}$ that maximizes $\text{var}(\mathbf{a}_1' \mathbf{X})$ subject to $\mathbf{a}_1' \mathbf{a}_1 = 1$. The second principal component is the linear combination $\mathbf{a}_2' \mathbf{X}$ that maximizes $\text{var}(\mathbf{a}_2' \mathbf{X})$ subject to $\mathbf{a}_2' \mathbf{a}_2 = 1$ and $\text{cov}(\mathbf{a}_1' \mathbf{X}, \mathbf{a}_2' \mathbf{X}) = 0$. At the i th step, the i th principal component is the linear combination $\mathbf{a}_i' \mathbf{X}$ that maximizes $\text{var}(\mathbf{a}_i' \mathbf{X})$ subject to $\mathbf{a}_i' \mathbf{a}_i = 1$ and $\text{cov}(\mathbf{a}_i' \mathbf{X}, \mathbf{a}_k' \mathbf{X}) = 0$ for $k < i$.

Let $\mathbf{X}' = [X_1, X_2, \dots, X_p]$ have the eigenvalue-eigenvector pairs $(\lambda_1, \mathbf{e}_1), (\lambda_2, \mathbf{e}_2), \dots, (\lambda_p, \mathbf{e}_p)$ with $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$. Let $Y_1 = \mathbf{e}_1' \mathbf{X}$, $Y_2 = \mathbf{e}_2' \mathbf{X}, \dots, Y_p = \mathbf{e}_p' \mathbf{X}$ be the principal components. Then

$$\sigma_{11} + \sigma_{22} + \dots + \sigma_{pp} = \sum_{i=1}^p \text{var}(X_i) = \lambda_1 + \lambda_2 + \dots + \lambda_p = \sum_{i=1}^p \text{var}(Y_i). \quad (10.2.1)$$

The proof is given in the appendix. Thus

$$\sum_{i=1}^p \text{var}(X_i) = \text{tr}(\Sigma) = \sum_{i=1}^p \text{var}(Y_i).$$

Result (10.2.1) says that

$$\text{Total population variance} = \sigma_{11} + \sigma_{22} + \dots + \sigma_{pp} = \lambda_1 + \lambda_2 + \dots + \lambda_p$$

and consequently, the proportion of total variance explained by the k th principal component is given by

$$\frac{\lambda_k}{\lambda_1 + \lambda_2 + \dots + \lambda_p}, \quad \text{for } k = 1, 2, \dots, p$$

If most (for instance, 80% to 90%) of the total variance, for large p , can be attributed to the first one, two, or three components, then these components can replace the original p variables without much loss of information (Linting et al; 2007).

10.3 The Permutation Approach

The permutation approach involves randomly and independently permuting the elements within the columns of the data matrix. If the variables are assumed to be interchangeable on the assumption of shared marginal distributions between variables, the data may be fully permuted between rows as well as columns. However, this assumption is unrealistic in practise because variables mostly differ in content and scaling. Therefore, usually the data are only permuted within the columns of the data set \mathbf{X} , on the assumption of shared marginal distributions between the objects (Good; 2000). Thus the total number of possible permuted data sets is n^{m-1} , where n represents the total number of observations and m is the number of permuted variables. Since this number increases rapidly with the number of objects and variables, usually a random sample of the total set of permutations is used. Linting et al. (2011) suggest using 999 permutations because if too few permutations are used, the p-value will be relatively large.

Suppose ν different permuted data matrices \mathbf{X}^* are constructed and analysed. This results in ν sets of eigenvalues $\{\lambda_p^*\}$, $p = 1, \dots, P$, with P the dimensionality of the PCA solution.

These eigenvalues are compared with the set of eigenvalues, $\{\lambda_p\}, p=1, \dots, P$, from the analysis of the original, unpermuted data matrix X . To test the significance of the p th eigenvalue of X , the exceedance probability (p-value) is computed. This is done by calculating the proportion of the values in the permutation distribution that is equal to or exceeds the observed statistic is computed. The p-value is computed as

$$p\text{-value} = P(\lambda_p^* \geq \lambda_p) = \frac{\text{number}(\lambda_p^* \geq \lambda_p)}{P+1}, \quad (10.3.1)$$

where the numerator represents the number of times a statistic from the permutation distribution is greater than or equal to the observed statistic and P is the number of permutations. (Buja and Eyuboglu; 1992). Under the null hypothesis, the observed data are assumed to be just another permutation of a random data set, thus the denominator in (10.3.1) is $P+1$ rather than P .

10.4 Two Permutation Strategies

Two different permutation strategies are considered. The first involves permuting the variables independently and concurrently and was proposed by Buja and Eyuboglu (1992) to establish the significance of the eigenvalues. The second method was proposed by Linting et al. (2011) and it involves permuting the variables independently and sequentially, that is, permuting one variable at a time, while keeping the others fixed.

The first form relates to the variance accounted for in the entire data set by the first c principal components, with c the number of components selected to represent the data set sufficiently. The total variance accounted for is equal to the sum of the eigenvalues of the first c components. Permuting all the variables in a data set concurrently enables the fit of a variable in an observed data set to be compared to the fit of variables with the same univariate distributions (its permutations) in a dataset with a completely random structure. Studies done by Buja and Eyuboglu (1992) show that this method is not appropriate to establish the significance of the contribution of a single variable to the principal component structure. Thus a more appropriate procedure is the use of the second method which assesses the significance of the variance accounted for of a variance, given the structure among the other variables.

The second permutation method relates to the contribution of each separate variable to the total variance accounted for. For this procedure, the elements x_{ij} of the columns X_j of the matrix are permuted independently. The philosophy of this permutation test is to destroy the pairing of observations, that is, to disconnect the link between the value of x_{ij} and x_{ik} that exists because they share the first index. To attain this, only the elements of one variable at a time should be permuted, while keeping the other variables fixed. This correspondence between the elements in the row of X is lost when the row elements of the columns are permuted. Permuting whole rows or columns has no effect (Buja and Eyuboglu; 1992). In the context of PCA the high dimensional hyper-ellipse that contains the row points in X is transformed into approximately a hypersphere by repeated pair wise switches of the row point positions along the coordinate axes. The hyper-ellipse's longest axes will most likely be shortened by the permutation of the data matrix. This corresponds to the idea that all eigenvalues are equal. Another way of interpreting the permutation of X is that the correlation structure of $X'X$ is destroyed. The restriction to standardised PCAs is not a necessary one as the variance/covariance structure (for centered data) or the sum of squares and cross product structure (for raw data) in $X'X$ will be destroyed too. A consequence of this approach is that more permutations are needed: If the first strategy is performed with 999 permutations, the alternative strategy involves $999 \times m$ permutations (with m the number of variables in the data set).

10.5 Numerical Example

Consider table 12.1 which represents the IAAF national track results for the 1984 Los Angeles Olympics. Table 12.1 shows the results of the 100m per second, 200m per second, 400m per second, 800m per minute, 1,500m per minute, 5,000m per minute, 10,000m per minute and the marathon per minute.

Table 10.1 The IAAF National Track Records for Men

| Country | 100 m (s) | 200 m (s) | 400 m (s) | 800 m (min) | 1,500 m (min) | 5,000 m (min) | 10,000 m (min) | Marathon (min) |
|-----------|-----------|-----------|-----------|-------------|---------------|---------------|----------------|----------------|
| Argentina | 10.39 | 20.81 | 46.84 | 1.81 | 3.70 | 14.40 | 26.36 | 137.72 |
| Australia | 10.31 | 20.06 | 44.84 | 1.74 | 3.57 | 13.28 | 27.66 | 128.30 |
| Austria | 10.44 | 20.81 | 46.82 | 1.79 | 3.60 | 13.26 | 27.72 | 135.90 |
| Belgium | 10.34 | 20.68 | 45.04 | 1.73 | 3.60 | 13.22 | 27.45 | 129.95 |

| | | | | | | | | |
|---------------|-------|-------|-------|------|------|-------|-------|--------|
| . | . | . | . | . | . | . | . | . |
| . | . | . | . | . | . | . | . | . |
| . | . | . | . | . | . | . | . | . |
| Switzerland | 10.37 | 20.46 | 45.78 | 1.78 | 3.55 | 13.22 | 27.91 | 131.20 |
| Taipei | 10.59 | 21.29 | 46.80 | 1.79 | 3.77 | 14.07 | 30.07 | 139.27 |
| Thailand | 10.39 | 21.09 | 47.91 | 1.83 | 3.84 | 15.23 | 32.65 | 159.90 |
| Turkey | 10.71 | 21.43 | 47.60 | 1.79 | 3.67 | 13.56 | 28.58 | 131.50 |
| USA | 9.93 | 19.75 | 43.86 | 1.73 | 3.53 | 13.2 | 27.43 | 128.22 |
| USSR | 10.07 | 20.00 | 44.60 | 1.75 | 3.59 | 13.2 | 27.53 | 130.55 |
| Western Samoa | 10.82 | 21.86 | 49.00 | 2.02 | 4.29 | 16.28 | 34.71 | 161.83 |

Source: Johnson, R.A. and Wichern, D.W. (2002). *Applied Multivariate Statistical Analysis*. 5th edition. Prentice Hall

Table 10.2 and 10.3 show the results from the PCA on the IAAF National track records for men. Looking at the eigenvalues, the first two principal components explain 86,88% of the sample variation. Thus the first two principal components may summarize the total sample variance.

Table 10.2 Eigenvalues of the Correlation Matrix for the Track Results

| Eigenvalues of the Correlation Matrix | | | | |
|---------------------------------------|-------------|------------|------------|------------|
| | λ_p | Difference | Proportion | Cumulative |
| 1 | 6.01395 | 5.07753 | 0.7517 | 0.7517 |
| 2 | 0.93642 | 0.35575 | 0.1171 | 0.8688 |
| 3 | 0.58067 | 0.43238 | 0.0726 | 0.9414 |
| 4 | 0.14829 | 0.00996 | 0.0185 | 0.9599 |
| 5 | 0.13833 | 0.06213 | 0.0173 | 0.9772 |
| 6 | 0.07620 | 0.01252 | 0.0095 | 0.9867 |
| 7 | 0.06369 | 0.02125 | 0.008 | 0.9947 |
| 8 | 0.04244 | | 0.0053 | 1 |

Table 10.3 Eigenvectors of the Correlation Matrix for the Track Results

| Eigenvectors | | | | | | | | |
|--------------|---------|----------|----------|----------|----------|----------|----------|----------|
| | Prin1 | Prin2 | Prin3 | Prin4 | Prin5 | Prin6 | Prin7 | Prin8 |
| m100 | 0.13679 | 0.94704 | 0.28507 | -0.04810 | -0.01842 | 0.00673 | 0.02103 | -0.00073 |
| m200 | 0.33992 | 0.15902 | -0.60353 | 0.62279 | -0.27093 | 0.15267 | -0.04952 | 0.08883 |
| m400 | 0.36596 | 0.07613 | -0.40926 | -0.22659 | 0.75472 | -0.26576 | -0.03270 | 0.01176 |
| m800 | 0.38495 | -0.04350 | -0.18548 | -0.57211 | -0.24104 | 0.56527 | 0.26181 | -0.20504 |
| m1500 | 0.39072 | -0.08484 | 0.03679 | -0.30718 | -0.44918 | -0.51795 | -0.13202 | 0.50691 |

| | | | | | | | | |
|----------|---------|----------|---------|---------|----------|----------|----------|----------|
| m5000 | 0.38755 | -0.11651 | 0.26169 | 0.06503 | -0.03006 | 0.02389 | -0.72350 | -0.48845 |
| m10000 | 0.38394 | -0.13625 | 0.27351 | 0.27101 | -0.03633 | -0.36027 | 0.61954 | -0.41332 |
| marathon | 0.36519 | -0.17552 | 0.45601 | 0.24318 | 0.30753 | 0.43342 | 0.05271 | 0.53249 |

To validate the PCA results, a permutation principal component analysis is performed. 500×8 different permuted data matrices are constructed and analysed using the second method described in section 10.4, that is, permuting only one column at a time, while keeping the other variables fixed.

A useful aid for determining an appropriate number of principal components is a scree plot. With the eigenvalues ordered from largest to smallest, a scree plot graphically illustrates the variance accounted for by each eigenvalue on the vertical axis versus the dimension on the horizontal axis. Figure 10.1 illustrates the variance accounted for, for a random sample of 30 permuted data matrices. Looking at the elbow bend in the scree plot, it can be seen that most of the variation can be explained by the first two principal components. The remaining eigenvalues are relatively small and all about the same size, so the first two sample principal components may effectively summarize the total sample variance.

Figure 10.1 *The Proportion of the Variance Accounted For per Dimension of the Principal Components of the Track Results.*

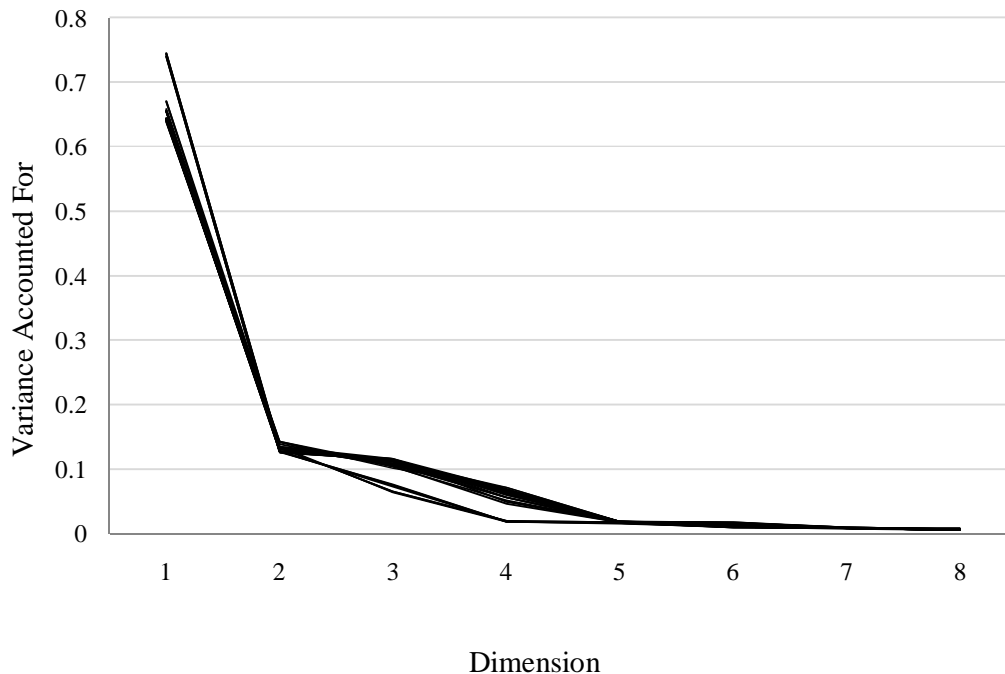


Table 10.4 presents the eight eigenvalues and the probabilities $P(\lambda_p^* \geq \lambda_p)$ for the eigenvalues $\lambda_p, p = 1, \dots, 8$. Table 10.4 gives clear evidence of the significant two-dimensional nature of the track results. The first principal component explains 60.14% of the total sample variance. The first two principal components, collectively, explain 86.88% of the total sample variance. Consequently, sample variation is summarised very well by two principal components. The p-values from the permuted eigenvalues clearly support the PCA results obtained, that is, the reduction of the 55 observations from eight variables to two variables as $\lambda_p = 0$ for $p = 1, 2$ and λ_p for $p = 3, 4, 5, 6, 7, 8$ are close to one.

Table 10.4 Empirical Eigenvalues λ_p and Permutation Test Results $P(\lambda_p^* \geq \lambda_p)$ for the Track Results.

| Dimension | Track Results | |
|-----------|---------------|---------------------------------|
| | λ_p | $P(\lambda_p^* \geq \lambda_p)$ |
| 1 | 6.01395 | 0.00000 |
| 2 | 0.93642 | 0.00000 |
| 3 | 0.58067 | 0.88745 |
| 4 | 0.14829 | 0.98317 |
| 5 | 0.13833 | 1.00000 |
| 6 | 0.07620 | 1.00000 |
| 7 | 0.06369 | 1.00000 |
| 8 | 0.04244 | 1.00000 |

11 Appendix

Proof of (9.6.1):

Consider

$$\varepsilon = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = (\mathbf{I}_n - \mathbf{H})$$

where $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Since \mathbf{H} is the projection matrix to the subspace spanned by the columns of \mathbf{X} , \mathbf{H} is an idempotent matrix for any design matrix \mathbf{X} , that is $\mathbf{H}^2 = \mathbf{H}$. Similarly, $\mathbf{I} - \mathbf{H}$ is also idempotent. Using the idempotent property of \mathbf{H} , the residual sum of squares under the full model can be written as

$$\varepsilon'\varepsilon = \mathbf{y}'(\mathbf{I} - \mathbf{H})^2\mathbf{y} = \mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y} = \mathbf{y}'(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})\mathbf{X}')\mathbf{y}$$

Similarly it can be shown that

$$\varepsilon^{KR1'}\varepsilon^{KR1} = \mathbf{y}'(\mathbf{I} - \mathbf{H}_1)^2\mathbf{y} = \mathbf{y}'(\mathbf{I} - \mathbf{H}_1)\mathbf{y} = \mathbf{y}'(\mathbf{I} - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)\mathbf{X}_1')\mathbf{y} \quad (11.8.1)$$

Thus

$$\mathbf{y}'(\mathbf{I} - \mathbf{H}_1)^2\mathbf{y} - \mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y} = \mathbf{y}'(\mathbf{X}(\mathbf{X}'\mathbf{X})\mathbf{X} - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)\mathbf{X}_1')\mathbf{y} \quad (11.8.2)$$

Thus equations (11.8.1) and (11.8.2) are similar.

Proof of (9.6.4):

Consider

$$F^{KR3} = \frac{\mathbf{y}'\left[\mathbf{X}^{KR1}(\mathbf{X}^{KR1'}\mathbf{X}^{KR1})^{-1}\mathbf{X}^{KR1'}\right]\mathbf{y}}{\mathbf{y}'\left[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right]\mathbf{y}} = \frac{\mathbf{y}'\mathbf{H}^{KR1}\mathbf{y}}{\mathbf{y}'[\mathbf{I} - \mathbf{H}]\mathbf{y}},$$

$$F^{KR4} = \frac{\mathbf{y}^{KR1} \left[\mathbf{X}^{KR1} (\mathbf{X}^{KR1} \mathbf{X}^{KR1})^{-1} \mathbf{X}^{KR1} \right] \mathbf{y}^{KR1}}{\mathbf{y}^{KR1} \left[\mathbf{I} - \mathbf{X}^{KR1} (\mathbf{X}^{KR1} \mathbf{X}^{KR1})^{-1} \mathbf{X}^{KR1} \right] \mathbf{y}^{KR1}} = \frac{\mathbf{y}^{KR1} \mathbf{H}^{KR1} \mathbf{y}^{KR1}}{\mathbf{y}^{KR1} [\mathbf{I} - \mathbf{H}^{KR1}] \mathbf{y}^{KR1}},$$

$$F^{KR1} = \frac{\mathbf{y}^{KR1} \left[\mathbf{X}^{KR1} (\mathbf{X}^{KR1} \mathbf{X}^{KR1})^{-1} \mathbf{X}^{KR1} \right] \mathbf{y}^{KR1}}{\mathbf{y}^{KR1} \left[\mathbf{I} - \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \right] \mathbf{y}^{KR1}} = \frac{\mathbf{y}^{KR1} \mathbf{H}^{KR1} \mathbf{y}^{KR1}}{\mathbf{y}^{KR1} [\mathbf{I} - \mathbf{H}] \mathbf{y}^{KR1}},$$

and

$$F^{KR2} = \frac{\mathbf{y}^{KR2} \left[\mathbf{X}^{KR2} (\mathbf{X}^{KR2} \mathbf{X}^{KR2})^{-1} \mathbf{X}^{KR2} \right] \mathbf{y}^{KR2}}{\mathbf{y}^{KR2} \left[\mathbf{I} - \mathbf{X}^{KR2} (\mathbf{X}^{KR2} \mathbf{X}^{KR2})^{-1} \mathbf{X}^{KR2} \right] \mathbf{y}^{KR2}} = \frac{\mathbf{y}^{KR2} \mathbf{H}^{KR2} \mathbf{y}^{KR2}}{\mathbf{y}^{KR2} [\mathbf{I} - \mathbf{H}^{KR2}] \mathbf{y}^{KR2}}$$

where

$$\mathbf{H} = \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}', \mathbf{H}_1 = \mathbf{X}_1 (\mathbf{X}_1' \mathbf{X}_1)^{-1} \mathbf{X}_1' \text{ and } \mathbf{H}^{KR1} = \mathbf{X}^{KR1} (\mathbf{X}^{KR1} \mathbf{X}^{KR1})^{-1} \mathbf{X}^{KR1},$$

Since $\check{\beta}_2 = \beta_2^{KR1} = \beta_2^{KR2}$, the only ingredient left to make an exact permutation test is to show that $F^{KR1} = F^{KR3}$, $F^{KR4} = F^{KR1}$ and $F^{KR2} = F^{KR4}$. In order to show that $F^{KR1} = F^{KR3}$, note that

$$\begin{aligned} \mathbf{H}_1 \mathbf{H}^{KR1} &= (\mathbf{X}_1 (\mathbf{X}_1' \mathbf{X}_1)^{-1} \mathbf{X}_1') (\mathbf{X}^{KR1} (\mathbf{X}^{KR1} \mathbf{X}^{KR1})^{-1} \mathbf{X}^{KR1}) \\ &= \mathbf{X}_1 (\mathbf{X}_1' \mathbf{X}_1)^{-1} (\mathbf{X}_1' \mathbf{X}^{KR1}) (\mathbf{X}^{KR1} \mathbf{X}^{KR1})^{-1} \mathbf{X}^{KR1}, \\ &= \mathbf{0} \end{aligned}$$

Using the equality above, it can be shown that the two numerators in F^{KR1} and F^{KR3} are equal:

$$\mathbf{y}^{KR1} \mathbf{H}^{KR1} \mathbf{y} = \mathbf{y}^{KR1} (\mathbf{I} - \mathbf{H}_1) \mathbf{H}^{KR1} (\mathbf{I} - \mathbf{H}_1) \mathbf{y} = \mathbf{y}' \mathbf{H}^{KR1} \mathbf{y}$$

Similarly, by using the fact that $\mathbf{H} = \mathbf{H}_1 + \mathbf{H}^{KR1}$, it can be shown that the denominators in F^{KR1} and F^{KR3} are equal:

$$\begin{aligned} \mathbf{y}^{KR1} (\mathbf{I} - \mathbf{H}) \mathbf{y}^{KR1} &= \mathbf{y}^{KR1} (\mathbf{I} - \mathbf{H}_1 - \mathbf{H}^{KR1}) \mathbf{y}^{KR1} \\ &= \mathbf{y}^{KR1} (\mathbf{I} - \mathbf{H}_1) ((\mathbf{I} - \mathbf{H}_1) - \mathbf{H}^{KR1}) (\mathbf{I} - \mathbf{H}_1) \mathbf{y} \\ &= \mathbf{y}' (\mathbf{I} - \mathbf{H}_1 - (\mathbf{I} - \mathbf{H}_1) \mathbf{H}^{KR1} (\mathbf{I} - \mathbf{H}_1)) \mathbf{y} \\ &= \mathbf{y}' (\mathbf{I} - \mathbf{H}_1 - \mathbf{H}^\perp) \mathbf{y} \\ &= \mathbf{y}' (\mathbf{I} - \mathbf{H}) \mathbf{y} \end{aligned}$$

Therefore

$$F^{\perp'} = \frac{\mathbf{y}' \mathbf{H}^{\perp} \mathbf{y}}{\mathbf{y}'(\mathbf{I} - \mathbf{H}) \mathbf{y}} = \frac{\mathbf{y}^{\perp'} \mathbf{H}^{\perp} \mathbf{y}^{\perp}}{\mathbf{y}^{\perp'}(\mathbf{I} - \mathbf{H}) \mathbf{y}^{\perp}} = F^{\ddagger}.$$

In order to show that $F^{KR4} = F^{KR1}$, it can be seen that the numerators of F^{KR4} and F^{KR1} are equal. It thus remains to show that their denominators are equal:

$$\begin{aligned} \mathbf{y}^{KR1'}(\mathbf{I} - \mathbf{H}_1)\mathbf{y}^{KR1} &= \mathbf{y}'(\mathbf{I} - \mathbf{H}_1)(\mathbf{I} - \mathbf{H}^{KR1})(\mathbf{I} - \mathbf{H}_1)\mathbf{y} \\ &= \mathbf{y}'((\mathbf{I} - \mathbf{H}_1) - (\mathbf{I} - \mathbf{H}_1)\mathbf{H}^{KR1}(\mathbf{I} - \mathbf{H}_1))\mathbf{y} \\ &= \mathbf{y}'(\mathbf{I} - \mathbf{H}_1 - \mathbf{H}^{KR1})\mathbf{y} \\ &= \mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y} \\ &= \mathbf{y}^{KR1'}(\mathbf{I} - \mathbf{H})\mathbf{y}^{KR1} \end{aligned}$$

Therefore

$$F^{KR4} = \frac{\mathbf{y}^{KR1'} \mathbf{H}^{KR1} \mathbf{y}^{KR1}}{\mathbf{y}^{KR1'}(\mathbf{I} - \mathbf{H}^{KR1})\mathbf{y}^{KR1}} = \frac{\mathbf{y}^{KR1'} \mathbf{H}^{KR1} \mathbf{y}^{KR1}}{\mathbf{y}^{KR1'}(\mathbf{I} - \mathbf{H})\mathbf{y}^{KR1}} = F^{KR1}.$$

In order to show that $F^{KR2} = F^{KR1}$, one can use the fact that $\mathbf{V}\mathbf{V}' = \mathbf{I} - \mathbf{H}_1$ to show that

$$\begin{aligned} \mathbf{y}^{KR2'} \mathbf{X}^{KR2} &= \mathbf{y}^{KR1'} \mathbf{V}\mathbf{V}' \mathbf{X}^{KR1} = \mathbf{y}^{KR1'}(\mathbf{I} - \mathbf{H}^{KR1})\mathbf{X}^{KR1} = \mathbf{y}^{KR1'} \mathbf{X}^{KR1}, \\ \mathbf{X}^{KR2'} \mathbf{X}^{KR2} &= \mathbf{X}^{KR1'} \mathbf{V}\mathbf{V}' \mathbf{X}^{KR1} = \mathbf{X}^{KR1'}(\mathbf{I} - \mathbf{H}^{KR1})\mathbf{X}^{KR1} = \mathbf{X}^{KR1'} \mathbf{X}^{KR1} \end{aligned}$$

and

$$\mathbf{y}^{KR2'} \mathbf{y}^{KR2} = \mathbf{y}^{KR1'} \mathbf{V}\mathbf{V}' \mathbf{y}^{KR1} = \mathbf{y}^{KR1'}(\mathbf{I} - \mathbf{H}^{KR1})\mathbf{y}^{KR1} = \mathbf{y}^{KR1'} \mathbf{y}^{KR1}$$

Therefore

$$F^{KR2} = \frac{\mathbf{y}^{KR2'} \mathbf{H}^{KR2} \mathbf{y}^{KR2}}{\mathbf{y}^{KR2'}(\mathbf{I}_{n-p} - \mathbf{H}^{KR2})\mathbf{y}^{KR2}} = \frac{\mathbf{y}^{KR1'} \mathbf{H}^{KR1} \mathbf{y}^{KR1}}{\mathbf{y}^{KR1'}(\mathbf{I} - \mathbf{H}^{KR1})\mathbf{y}^{KR1}} = F^{KR1}.$$

Proof of 10.2.1

Since $\sigma_{11} + \sigma_{22} + \dots + \sigma_{pp} = \text{tr}(\Sigma)$ one can write $\Sigma = \mathbf{E}\Lambda\mathbf{E}$ where Σ is the covariance matrix, Λ is the diagonal matrix of eigenvalues and $\mathbf{E} = [\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_p]$ so that

$\mathbf{E}\mathbf{E}' = \mathbf{E}'\mathbf{E} = \mathbf{I}_p$. Thus

$$\text{tr}(\Sigma) = \text{tr}(\mathbf{E}\Lambda\mathbf{E}') = \text{tr}(\Lambda\mathbf{E}'\mathbf{E}) = \text{tr}(\Lambda) = \lambda_1 + \lambda_2 + \dots + \lambda_p$$

Thus

$$\sum_{i=1}^p \text{var}(X_i) = \text{tr}(\Sigma) = \text{tr}(\Lambda) = \sum_{i=1}^p \text{var}(Y_i)$$

12 References

- Anderson MJ, A new method for non-parametric multivariate Analysis of Variance. *Aus. J. Ecol.* 1999, 26:32-46.
- Anderson MJ. A new method for non-parametric multivariate Analysis of Variance. *Aus. J. Ecol.* 2001, 26:32-46.
- Anderson MJ, Ter Braak CJF. Permutation tests for multifactorial analysis of variance, *Journal of Statistical Computation and Simulation* 2003, 73:85-113.
- Bailer AJ. Testing variance equality with randomization tests. *J Stat Comput Simul* 1989, 31:168.
- Berry KJ, Mielke PW. Moment approximations as an alternative to the F test in analysis of variance. *Br J Math Stat Psychol* 1983, 36:2026206.
- Berry KJ, Mielke PW. Moment approximations as an alternative to the F test in analysis of variance. *Br J Math Stat Psychol* 1983, 36:2026206.
- Berry KJ, Mielke PW. Goodman and Kruskal's tau-b statistic: a nonasymptotic test of significance. *Sociol Methods Res* 1985, 13:5436550.
- Berry KJ, Mielke PW. Computation of exact and approximate probability values for a matched-pairs permutation test. *Commun Stat Simul Comput* 1985, 14:2296248.
- Berry KJ, Mielke PW. Exact chi-square and Fisher's exact probability test for 3 by 2 cross-classification tables. *Educ Psychol Meas* 1987, 47:6316636.
- Berry KJ, Mielke PW, Helmericks SG. Exact confidence limits for proportions. *Educ Psychol Meas* 1988, 48:7136716.
- Berry KJ, Mielke PW. A FORTRAN program for permutation covariate analyses of residuals based on Euclidean distance. *Psychol Rep* 1998, 82:3716375.
- Berry KJ, Mielke PW, Helmericks SG. An algorithm to generate discrete probability distributions: binomial, hypergeometric, negative binomial, inverse hypergeometric, and Poisson. *Behav Res Methods Instrum Comput* 1994, 26:3666367.
- Berry KJ, Mielke PW. Exact cumulative probabilities for the multinomial distribution. *Educ Psychol Meas* 1995, 55:7696772.
- Berry KJ, Mielke PW. Least sum of absolute deviations regression: distance, leverage, and influence. *Percept Mot Skills* 1998, 86:106361070.
- Berry KJ, Mielke PW. Least absolute regression residuals: analyses of randomized designs. *Psychol Rep* 1999, 84:9476954.
- Berry KJ, Mielke PW, Mielke HW. The Fisher-Pitman permutation test: an attractive alternative to the F test. *Psychol Rep* 2002, 90:4956502.
- Berry KJ, Johnston JE, Mielke PW. *Permutation Methods*. John Wiley & Sons; 2011.
- Buja A., Eyuboglu N. Remarks on parallel analysis. *Multivariate Behavioral Research*, 27,509-540.

- Cade BS, Richards JD. Permutation tests for least absolute deviation regression. *Biometrics* 1996, 52:886-902.
- Chen RU, Dunlop WP. SAS Procedures for approximate randomization tests. *Behavior Research Methods, Instruments, & Computers* 1993, 25 (3). 406-409.
- Dallal GE. PITMAN: a FORTRAN program for exact randomization tests. *Comput Biomed Res* 1988, 21:9615.
- Dwass M. Modified randomization tests for nonparametric hypotheses. *Ann Math Stat* 1957, 28:1816187
- Eden T, Yates F. On the validity of Fisher's z test when applied to an actual example of non-normal data. *J Agric Sci* 1933, 23:6617.
- Edgington ES. *Randomization Tests*. New York: Marcel Dekker; 1980.
- Edgington ES. *Randomization Tests*. 2nd ed. New York:Marcel Dekker; 1987.
- Edgington ES. *Randomization Tests*.3rd ed. New York: Marcel Dekker; 1995.
- Edgington ES, Onghena P. *Randomization Tests*. 4th ed. Boca Raton, FL: Chapman&Hall/CRC; 2007.
- Efron, B. And Tibshirini, RJ. *An Introduction to the Bootstrap*. Chapman & Hall, London; 1993.
- Fisher RA. The logic of inductive inference. *J R Stat Soc* 1935, 98:39682.
- Fisher RA. *The Design of Experiments*. Edinburgh: Oliver and Boyd; 1935.
- Fisher RA. The coefficient of racial likeness and the future of craniometry. *J R Anthropol Inst* 1936, 66:57663.
- Freedman D, Lane D. (1983). A nonstochastic interpretation of reported significance levels. *Journal of Business and Economic Statistics* 1983, 1:2926298.
- Geary RC. Some properties of correlation and regression in a limited universe. *Metron Riv Int Stat* 1927, 7:836119.
- Gonzales L, Manly BFJ. Analysis of variance by randomization with small data sets. *Environmetrics* 1998, 9:53-65.
- Good PI. *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses*. New York: Springer-Verlag; 1994.
- Good PI. *Permutation, Parametric and Bootstrap Tests of Hypotheses*. New York: Springer-Verlag; 1994.
- Good PI. *Resampling Methods: A Practical Guide to Data Analysis*. Boston, MA: Birkhauser; 1999.
- Good PI. *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses*. 2nd ed. New York: Springer-Verlag; 2000.

- Good PI. *Resampling Methods: A Practical Guide to Data Analysis*. 2nd ed. Boston, MA: Birkhauser; 2001.
- Good PI. *Permutation, Parametric and Bootstrap Tests of Hypotheses*. 3rd ed. New York: Springer-Verlag; 2005.
- Hoefding W. The large-sample power of tests based on permutations of observations. *Ann Math Stat* 1952, 23:1696192.
- Hotelling H, Pabst MR. Rank correlation and tests of significance involving no assumption of normality. *Ann Math Stat* 1936, 7:29643.
- Hubert L. *Assignment Methods in Combinatorial Data Analysis*. New York: Marcel Dekker; 1987.
- Huh M, Jhun M. Random permutation testing in multiple linear regression. *Communications in Statistics* 2001, 10:20236 2032.
- Jung BC, Jhun M, Song, SH. A new random permutation test in ANOVA models. *Statistical Papers* 2006, 48:47662.
- Johnson RA & Wichern DW *Applied Multivariate Statistical Analysis*. 5th ed. Prentice-Hall, NJ; 2002.
- Keller-McNulty S, Higgins JJ. Effect of tail weight and outliers and power and type-I error of robust permutation tests for location. *Commun Stat Comput Simul* 1987, 16:17635.
- Kempthorne O. The randomization theory of experimental inference. *J Am Stat Assoc* 1955, 50:9466967.
- Kendall MG, Kendall SFH, Babington Smith B. The distribution of Spearman's coefficient of rank correlation in a universe in which all rankings occur an equal number of times. *Biometrika* 1939, 30:2516273.
- Kennedy PE. Randomization tests in econometrics. *J Bus Econ Stat* 1995, 13:85694.
- Kruskal WH, Wallis WA. Use of ranks in one-criterion variance analysis. *J Am Stat Assoc* 1952, 47:5836621. Erratum in: *J Am Stat Assoc* 1953, 48:9076911.
- Lingting M, Meulman, JJ, Groenen PJF, van der Kooij AJ. Nonlinear principal component analysis: introduction and application. *Psychological Methods* 2007a, 12:336-358.
- Lingting M, Meulman, JJ, Groenen PJF, van der Kooij AJ. Stability of nonlinear principal components analysis: an empirical study using the balanced bootstrap. *Psychological Methods* 2007b, 12:359-378.
- Linting M, van Os BJ, Meulman JJ. Statistical significance of the contribution of variables to the PCA solution: an alternative permutation strategy. *Psychometrika* 2011, 3, 440-460.
- Ludbrook J. Advantages of permutation (randomization) tests in clinical and experimental pharmacology and physiology. *Clin Exp Pharmacol Physiol* 1994, 21:6736686.
- Manly BFJ. *Randomization and Monte Carlo Methods in Biology*. London: Chapman & Hall; 1991.
- Manly BFJ. *Randomization and Monte Carlo Methods in Biology*. 2nd ed. London: Chapman & Hall; 1997.

- Manly BFJ. *Randomization, Bootstrap and Monte Carlo Methods in Biology*. 3rd ed. Boca Raton, FL:Chapman & Hall/CRC; 2007.
- Mann HB, Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. *Ann Math Stat* 1947, 18:50660.
- McCarthy MD. On the application of the z -test to randomized blocks. *Ann Math Stat* 1939, 10:3376359.
- Mehta CR, Patel NR. A network algorithm for the exact treatment of the $2 \times k$ contingency table. *Commun Stat Simul Comput* 1980, 9:6496664.
- Mielke PW. Asymptotic behavior of two-sample tests based on powers of ranks for detecting scale and location alternatives. *J Am Stat Assoc* 1972, 67:8506854.
- Mielke PW, Berry KJ. An extended class of matched pairs tests based on powers of ranks. *Psychometrika* 1976, 41:896100.
- Mielke PW, Berry KJ, Johnson ES. Multi-response permutation procedures for a priori classifications. *Commun Stat Theory Methods* 1976, 5:140961424.
- Mielke PW. On asymptotic non-normality of null distributions of MRPP statistics. *Commun Stat Theory Methods* 1979, 8:154161550. Errata in: *Commun Stat Theory Methods* 1981, 10:1795 and 1982, 11:847.
- Mielke PW, Sen PK. On asymptotic non-normal null distributions for locally most powerful rank test statistics. *Commun Stat Theory Methods* 1981, 10:107961094.
- Mielke PW. Multiresponse permutation procedures. In: *Encyclopedia of Statistical Sciences*. Vol.V. New York: John Wiley & Sons, 1985, 7246727.
- Mielke PW, Berry KJ. Non-asymptotic inferences based on the chi-square statistic for r by c contingency tables. *J Stat Plann Infer* 1985, 12:41645.
- Mielke PW, Berry KJ. Permutation-based multivariate regression analysis: the case for least sum of absolute deviations regression. *Ann Oper Res* 1997, 74:2596268.
- Mielke PW, Berry KJ. *Permutation Methods: A Distance Function Approach*. New York: Springer-Verlag; 2001.
- Neyman J. Proba uzasadnienia zastosowa ń rachunku prawdopodobienstwa do doswiadczen polowych (On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Rocz Nauk Rolniczych* 1923, 10:1651. Translated by Dabrowska DM and Speed TP in *Stat Sci* 1990, 5:4656472.
- Neyman J, Pearson ES. On the use and interpretation of certain test criteria for purposes of statistical inference: part I. *Biometrika* 1928, 20A:1756240.
- Neyman J, Pearson ES. On the use and interpretation of certain test criteria for purposes of statistical inference: part II. *Biometrika* 1928, 20A:2636294.
- O'Reilly FJ, Mielke PW. Asymptotic normality of MRPP statistics from invariance principles of U statistics. *Commun Stat Theory Methods* 1980, 9:6296637.
- Pearson ES. Some aspects of the problem of randomization. *Biometrika* 1937, 29:53664.

- Pesarin F. *Multivariate Permutation Tests: With Applications in Biostatistics*. Chichester: John Wiley & Sons;2001.
- Pitman E.J.G. Significance tests which may be applied to samples from any populations. *Suppl J R Stat Soc* 1937, 4:1196130.
- Pitman E.J.G. Significance tests which may be applied to samples from any populations: II. The correlation coefficient test. *Suppl J R Stat Soc* 1937, 4:2256232.
- Pitman E.J.G. Significance tests which may be applied to samples from any populations: III. The analysis of variance test. *Biometrika* 1938, 29:3226335.
- Steyn AGW, Smit CF, Du Toit SHC, Strasheim C. *Modern Statistics in Practice*. Van Schaik; 2000.
- Still, AW, White, AP. (1981), The approximate randomization test as an alternative to the F-test in the analysis of variance. *British Journal of Mathematical and Statistical Psychology* 1981,34, 243-252.
- Tantawanich S. Permutation Test for a multiple Linear Regression model. University Sumutphrakhran Hong Kong 2006, 641-647.
- Ter Braak C.J.F. Permutation versus bootstrap significance tests in multiple regression and ANOVA, *Bootstrapping and Related Techniques*, Springer, Berlin 1992, 79-86.
- Tucker DF, Mielke PW, Reiter ER. The verification of numerical models with multivariate randomized block permutation procedures. *Meteorol Atmos Phys* 1989, 40:1816188.
- Welch BL. On the z-test in randomized blocks and Latin squares. *Biometrika* 1937, 29:21652.
- Welch BL. On tests for homogeneity. *Biometrika* 1938, 30:1496158.
- Welch WJ. Construction of permutation tests. *Journal of American Statistical Association* 1990, 85:693-698.
- Wilcoxon F. Individual comparisons by ranking methods. *Biometrics Bull* 1945, 1:80683.
- Wilcoxon F. Probability tables for individual comparisons by ranking methods. *Biometrics* 1947, 3:1196122.

13 Appendix

13.1 Code for Section 8.3

```
title'Number of ants eaten from june to september by small and large
lizards';
data lizard;
input size $ months $ ants;
datalines;
sjun 13
sjul 8
saug 515
ssep 18
sjun 242
sjul 59
saug 488
ssep 44
sjun 105
sjul 20
saug 88
ssep 21
ljun 182
ljul 24
laug 460
lsep 140
ljun 21
ljul 312
laug 1223
lsep 40
ljun 7
ljul 68
laug 990
lsep 27
;

/*Perform analysis of variance for balanced data*/

procanova data = lizard;
class size months;
model ants = size months size*months;
run;
quit;

prociml;
resethnolog;

a = 4;
b = 2;
n = 24;
m = 10;
simul = 10000;
reject = J(1,1,0);

y_sw = J(n,1,0);
y_1_1 = J(6,1,0);
```

```

y_1_2 = J(6,1,0);
y_1_3 = J(6,1,0);
y_1_4 = J(6,1,0);
y_2_1 = J(12,1,0);
y_2_2 = J(12,1,0);

y_1_1_sw = J(6,1,0);
y_1_2_sw = J(6,1,0);
y_1_3_sw = J(6,1,0);
y_1_4_sw = J(6,1,0);
y_2_1_sw = J(12,1,0);
y_2_2_sw = J(12,1,0);

y_1_1_b = J(6,1,0);
y_1_2_b = J(6,1,0);
y_1_3_b = J(6,1,0);
y_1_4_b = J(6,1,0);
y_2_1_b = J(12,1,0);
y_2_2_b = J(12,1,0);

y_perm = J(24,1,0);
y_perm_sw = J(24,1,0);
y_perm_b = J(24,1,0);
ybar_Si = J(2,1,0);
ybar_M = J(5,1,0);
ybar_Si_sw = J(2,1,0);
ybar_Si_b = J(2,1,0);
ybar_M_sw = J(5,1,0);
y_ijo = J(24,1,0);
y_ijo_M = J(24,1,0);
fa = J(1,1,0);
fb = J(1,1,0);
F_Manly = J(simul,1,0);
F_sw = J(simul,1,0);
F_b = J(simul,1,0);
p_value_m = J(simul,1,0);
p_value_sw = J(simul,1,0);
p_value_b = J(simul,1,0);
p_value_j = J(simul,1,0);
count = 0;

/* The Permutation Approach */

y = { 13, 8, 515, 18, 242, 59, 488, 44, 105, 20, 88, 21, 182, 24, 460, 140,
21, 312, 1223, 40, 7, 68, 990, 27};
do j = 1 to simul;
seed = 1234567;
do i = 1 to 24;
call ranperm(seed,m,y);
y_perm = y;
end;

y_1_1[1] = y_perm[1];
y_1_1[2] = y_perm[1+4];
y_1_1[3] = y_perm[5+4];
y_1_1[4] = y_perm[9+4];
y_1_1[5] = y_perm[13+4];
y_1_1[6] = y_perm[17+4];

```

```

y_1_2[1] = y_perm[2];
y_1_2[2] = y_perm[2+4];
y_1_2[3] = y_perm[6+4];
y_1_2[4] = y_perm[10+4];
y_1_2[5] = y_perm[14+4];
y_1_2[6] = y_perm[18+4];

y_1_3[1] = y_perm[3];
y_1_3[2] = y_perm[3+4];
y_1_3[3] = y_perm[7+4];
y_1_3[4] = y_perm[11+4];
y_1_3[5] = y_perm[15+4];
y_1_3[6] = y_perm[19+4];

y_1_4[1] = y_perm[4];
y_1_4[2] = y_perm[4+4];
y_1_4[3] = y_perm[8+4];
y_1_4[4] = y_perm[12+4];
y_1_4[5] = y_perm[16+4];
y_1_4[6] = y_perm[20+4];

y_2_1[1:12] = y_perm[1:12];
y_2_2[1:12] = y_perm[13:24];

ybar_ju = y_1_2[+]/(6*24);
ybar_jul= y_1_2[+]/(6*24);
ybar_au = y_1_3[+]/(6*24);
ybar_se = y_1_4[+]/(6*24);
ybar_s= y_2_1[+]/(2*24);
ybar_l= y_2_2[+]/(2*24);

ybar_M[1] = ybar_ju[1];
ybar_M[2] = ybar_jul[1];
ybar_M[3] = ybar_au[1];
ybar_M[4] = ybar_se[1];

ybar_Si[1] = ybar_s[1];
ybar_Si[2] = ybar_l[1];
y_ijo= y/24;
y_ooo = y[+]/n;

/* The Manly Method */

/* Calculate the permuted F Manly values */
do s = 1to a;
    do t = 1to b;
        do u = 1to n;
            fa = (y_ijo[u] - ybar_Si[t] - ybar_M[s] + y_ooo)**2
        /((2-1)*(5-1));
            fb = (y[u] - y_ijo[u])**2/(2*5*23);
        end;
    end;
end;
F_Manly[j] = fa/fb;
if F_Manly >= 3 then count = count + 1;
print fa fb ;

/*Calculate the permuted p value*/
p_value_m[j] = count / simul;

```

```

ifp_value_m[j] < 0.05 then reject = reject + 1;
  printp_value_m;
/* The Still and White Method */

/* Calculate the new y Still and White values */
doss = 1 to a;
  dott = 1 to b;
  douu = 1 to n;
  y_sw[uu] = y[uu] - ybar_Si[tt] - ybar_M[ss] +
y_ooo;
  end;
end;
end;
print y_sw;
seed = 1234567;
do i = 1 to 24;
  call ranperm(seed, m, y_sw);
  y_perm_sw = y_sw;
end;

y_1_1_sw[1] = y_perm_sw[1];
y_1_1_sw[2] = y_perm_sw[1+4];
y_1_1_sw[3] = y_perm_sw[5+4];
y_1_1_sw[4] = y_perm_sw[9+4];
y_1_1_sw[5] = y_perm_sw[13+4];
y_1_1_sw[6] = y_perm_sw[17+4];

y_1_2_sw[1] = y_perm_sw[2];
y_1_2_sw[2] = y_perm_sw[2+4];
y_1_2_sw[3] = y_perm_sw[6+4];
y_1_2_sw[4] = y_perm_sw[10+4];
y_1_2_sw[5] = y_perm_sw[14+4];
y_1_2_sw[6] = y_perm_sw[18+4];

y_1_3_sw[1] = y_perm_sw[3];
y_1_3_sw[2] = y_perm_sw[3+4];
y_1_3_sw[3] = y_perm_sw[7+4];
y_1_3_sw[4] = y_perm_sw[11+4];
y_1_3_sw[5] = y_perm_sw[15+4];
y_1_3_sw[6] = y_perm_sw[19+4];

y_1_4_sw[1] = y_perm_sw[4];
y_1_4_sw[2] = y_perm_sw[4+4];
y_1_4_sw[3] = y_perm_sw[8+4];
y_1_4_sw[4] = y_perm_sw[12+4];
y_1_4_sw[5] = y_perm_sw[16+4];
y_1_4_sw[6] = y_perm_sw[20+4];

y_2_1_sw[1:12] = y_perm_sw[1:12];
y_2_2_sw[1:12] = y_perm_sw[13:24];

ybar_ju_sw = y_1_2_sw[+]/(6*24);
ybar_jul_sw = y_1_2_sw[+]/(6*24);
ybar_au_sw = y_1_3_sw[+]/(6*24);
ybar_se_sw = y_1_4_sw[+]/(6*24);
ybar_s_sw = y_2_1_sw[+]/(2*24);
ybar_l_sw = y_2_2_sw[+]/(2*24);

ybar_M_sw[1] = ybar_ju_sw[1];

```



```

ybar_M_sw[2] = ybar_jul_sw[1];
ybar_M_sw[3] = ybar_au_sw[1];
ybar_M_sw[4] = ybar_se_sw[1];

ybar_Si_sw[1] = ybar_s_sw[1];
ybar_Si_sw[2] = ybar_l_sw[1];
y_ijo = y/24;
y_ooo = y_sw[+]/n;

/* Calculate the permuted F Still and White values */
dost = 1to a;
    dott = 1to b;
        dout = 1to n;
            fa = (y_ijo[ut] - ybar_Si[tt] - ybar_M[st] +
y_ooo)**2 / ((2-1)*(5-1));
            fb = (y[ut] - y_ijo[ut])**2/(2*5*23);
        end;
    end;
end;
F_sw[j] = fa/fb;
if F_sw >= 3 then count = count + 1;
printfafb ;

/*Calculate the permuted p value*/
p_value_sw[j] = count / simul;
if p_value_sw[j] < 0.05 then reject = reject + 1;

printp_value_sw;

/* The TerBraak Method */

y_b = J(n,1,0);
y_ijo_M = y/n;

/* Calculate the new y Ter Braak values */
y_b = y - y_ijo_M;

seed = 1234567;
do i = 1to24;
    call ranperm(seed,m,y_b);
    y_perm_sw = y_sw;
end;

y_ijo_M = y_b/24;
y_1_1_b[1] = y_perm_b[1];
y_1_1_b[2] = y_perm_b[1+4];
y_1_1_b[3] = y_perm_b[5+4];
y_1_1_b[4] = y_perm_b[9+4];
y_1_1_b[5] = y_perm_b[13+4];
y_1_1_b[6] = y_perm_b[17+4];

y_1_2_b[1] = y_perm_b[2];
y_1_2_b[2] = y_perm_b[2+4];
y_1_2_b[3] = y_perm_b[6+4];
y_1_2_b[4] = y_perm_b[10+4];
y_1_2_b[5] = y_perm_b[14+4];
y_1_2_b[6] = y_perm_b[18+4];

y_1_3_b[1] = y_perm_b[3];

```

```

y_1_3_b[2] = y_perm_b[3+4];
y_1_3_b[3] = y_perm_b[7+4];
y_1_3_b[4] = y_perm_b[11+4];
y_1_3_b[5] = y_perm_b[15+4];
y_1_3_b[6] = y_perm_b[19+4];

y_1_4_b[1] = y_perm_b[4];
y_1_4_b[2] = y_perm_b[4+4];
y_1_4_b[3] = y_perm_b[8+4];
y_1_4_b[4] = y_perm_b[12+4];
y_1_4_b[5] = y_perm_b[16+4];
y_1_4_b[6] = y_perm_b[20+4];

y_2_1_b[1:12] = y_perm_b[1:12];
y_2_2_b[1:12] = y_perm_b[13:24];

ybar_ju_b = y_1_2_b[+]/(6*24);
ybar_jul_b = y_1_2_b[+]/(6*24);
ybar_au_b = y_1_3_b[+]/(6*24);
ybar_se_b = y_1_4_b[+]/(6*24);
ybar_sb = y_2_1_b[+]/(2*24);
ybar_lb = y_2_2_b[+]/(2*24);

ybar_M_sw[1] = ybar_ju_b[1];
ybar_M_sw[2] = ybar_jul_b[1];
ybar_M_sw[3] = ybar_au_b[1];
ybar_M_sw[4] = ybar_se_b[1];

ybar_Si_b[1] = ybar_sb[1];
ybar_Si_b[2] = ybar_lb[1];
y_ijo = y/24;
y_ooo = y_b[+]/n;

/* Calculate the permuted F Ter Braak values */
dosp = 1to a;
    dotp = 1to b;
        do up = 1to n;
            fa = (y_ijo[up] - ybar_Si[tp] - ybar_M[sp] +
y_ooo)**2 / ((2-1)*(5-1));
            fb = (y[up] - y_ijo[up])**2 / (2*5*23);
            end;
        end;
    end;
    F_b[j] = fa/fb;
    if F_b >= 3 then count = count + 1;
    print fafb ;

p_value_b[j] = count / simul;
if p_value_b[j] < 0.05 then reject = reject + 1;
print p_value_b;

/* The Jung et al. Method */
/* set up design matrix */
X = {0 -1 -1, 0 -11, 01 -1, 0 11, 1 -1 -1, 1 -11,
11 -1, 111, 2 -1 -1, 2 -11, 21 -1,
211, 3 -1 -1, 3 -11, 31 -1, 311,
4 -1 -1, 4 -11, 41 -1, 411, 5 -1 -1,
5 -11, 51 -1, 511} ;
nfact = 999;

```

```

H = J(n,n,0);
VV = J(n,n,0);
XH = X[,2:2];
X0 = X[,3:3];
H = XH*inv(XH`*XH)*XH`;
VV = I(n) - H;
calleigen (eval,vec,VV);
V = J(n,n,1);
V = evec*(diag(eval)**0.5)*evec`;
y_v = V`*y;
X0_v = V*X0;
u = J(n,1,0);
y_v_perm = J(n,nfact,0);
F_perm_a = J(nfact,1,0);
F_perm_b = J(nfact,1,0);
F_perm = J(nfact,1,0);

/*Generate the permuted F values*/;
do h = 1:nfact;
    fa[h] =
(Y_v_perm[,h]`*(X0_v*inv(X0_v`*X0_v)*X0_v)`*Y_v_perm[,h])/2;
    fa[h] = (Y_v_perm[,h]`*(I(n)-
X*inv(X`*X)*X)`*Y_v_perm[,h])/(n-2-2);
    F_perm[h] = fa[h]/fa[h];
end;

/*Calculate the permuted p value*/
p_value_j[j] = count / simul;
if p_value_j[j] < 0.05 then reject = reject + 1;
end;

run;
quit;

```

13.2 Code for Section 8.5

```

title"Simulation of ANOVA procedures";
prociml;
resetnolog;

n=12;
nfact = 1000;
simul = 10000;

p_value = J(simul,1,0);
reject = 0;
a = 4;
b = 2;
n = 24;
m = 10;
simul = 10000;
reject = J(1,1,0);
y_perm = J(24,1,0);
y_perm_sw = J(24,1,0);
y_perm_b = J(24,1,0);
ybar_Si = J(2,1,0);
ybar_M = J(5,1,0);

```

```

ybar_Si_sw = J(2,1,0);
ybar_Si_b = J(2,1,0);
ybar_M_sw = J(5,1,0);
y_ijo = J(24,1,0);
y_ijo M = J(24,1,0);
fa = J(1,1,0);
fb = J(1,1,0);
F_Manly = J(simul,1,0);
F_sw = J(simul,1,0);
F_b= J(simul,1,0);
p_value_m = J(simul,1,0);
p_value_sw = J(simul,1,0);
p_value_b = J(simul,1,0);
p_value_j = J(simul,1,0);
count = 0;

/*For each i, xs are generated from a uniform distribution
Generate the reference sample*/;

X = { 0.43910171.92319383.98945840.6767025,
1.03150021.1409022.98764164.7927917,
0.48146352.41864482.42143153.3979654,
.... ,
.... ,
.... ,
0.88368211.18391314.49023045.1701252,
0.5371720.36115121.36273923.3279271,
0.72649341.66435261.19636460.6123876 } ;

/*Generate the errors from a normal, exponential, t and uniform
distribution*/
e = { -1.613007,
      -0.137757,
      -0.29423,
      . ,
      . ,
      . ,
      -0.711689,
      -1.00243,
      -1.242766 };

b1 = {1};
b2 = {1};
b3 = {0};
b4 = {0};
b = b1//b2//b3//b4;
b0 = b3//b4;

/*Generate the reference sample Y*/;
y = X*b + e;

/*Generate V and multiply to the reduced model*/
H = J(n,n,0);
VV = J(n,n,0);
XH = X[,1:2];
X0 = X[,3:4];
H = XH*inv(XH`*XH)*XH`;
VV = I(n) - H;
calleigen (eval,vec,VV);
V = J(n,n,1);

```

```

V = evec*(diag(eval)**0.5)*evec`;
y_v = V`*y;
X0_v = V*X0;
e_v = V`*e;
u = J(n,1,0);
e_perm = J(n,nfact,0);
y_v_perm = J(n,nfact,0);
F_perm_a = J(nfact,1,0);
F_perm_b = J(nfact,1,0);
F_perm = J(nfact,1,0);
Fa = J(1,1,0);
Fb = J(1,1,0);
F_ref = J(1,1,0);

do s = 1tosimul;
  /*Generate the e permutations*/
  do j = 1tonfact;
    do l = 1to n;
      u[l] = rannor(0);
    end;

    call SORTNDX( ndx, u, {1}, {1} );
    u = u[ndx,];

    do i = 1to n;
      e_perm[i,j] = e[ndx[i]];
    end;
  end;

  /*Generate the reference sample F;*/
  Fa = (y_v`*(X0_v*inv(X0_v`*X0_v)*X0_v`)*y_v)/2;
  Fb = (y_v`*(I(n)- X0_v*inv(X0_v`*X0_v)*X0_v`)*y_v)/(n-2-2);
  F_ref = Fa/Fb;

  /*Generate the permuted y values*/;
  do g = 1tonfact;
    y_v_perm[,g] = X0_v*b0 + e_perm[,g];
  end;

  /*Generate the permuted F values for Manly Test*/
  do s = 1to a;
    do t = 1to b;
      do u = 1to n;
        fa = (y_ijo[u] - ybar_Si[t] - ybar_M[s] + y_ooo)**2
      /((2-1)*(5-1));
        fb = (y[u] - y_ijo[u])**2/(2*5*23);
      end;
    end;
    F_Manly[j] = fa/fb;
    if F_Manly >= 17.4 then count = count + 1;

  /*Calculate the permuted p value*/
  p_value_m[j] = count / simul;
  if p_value_m[j] < 0.05 then reject = reject + 1;

  /* Calculate the new y Still snd White values */
  doss = 1to a;

```

```

        dott = 1to b;
        douu = 1to n;
        y_sw[uu] = y[uu] - ybar_Si[tt] - ybar_M[ss] +
y_ooo;
        end;
    end;
end;
print y_sw;
seed = 1234567;
do i = 1to24;
    call ranperm(seed,m,y_sw);
    y_perm_sw = y_sw;
end;

/* Calculate the permuted F Still and White values */
dost = 1to a;
    dott = 1to b;
    dout = 1to n;
    fa = (y_ijo[ut] - ybar_Si[tt] - ybar_M[st] +
y_ooo)**2 / ((2-1)*(5-1));
    fb = (y[ut] - y_ijo[ut])**2 / (2*5*23);
    end;
end;
F_sw[j] = fa/fb;
if F_sw >= 3 then count = count + 1;

/* Calculate the permuted p value for Still and White */
p_value_sw[j] = count / simul;
if p_value_sw[j] < 0.05 then reject = reject + 1;

/* Calculate the new y TerBraak values */
y_b = y - y_ijo_M;

/* Calculate the permuted F TerBraak values */
dosp = 1to a;
    dotp = 1to b;
    do up = 1to n;
        fa = (y_ijo[up] - ybar_Si[tp] - ybar_M[sp] +
y_ooo)**2 / ((2-1)*(5-1));
        fb = (y[up] - y_ijo[up])**2 / (2*5*23);
        end;
    end;
F_b[j] = fa/fb;
if F_b >= 3 then count = count + 1;

    p_value_b[j] = count / simul;
if p_value_b[j] < 0.05 then reject = reject + 1;

/* Generate the permuted F Jung et al. values */;
do h = 1to nfact;
    fa[h] =
(Y_v_perm[,h]`*(X0_v`*inv(X0_v`*X0_v`)*X0_v`)*Y_v_perm[,h])/2;
    fa[h] = (Y_v_perm[,h]`*(I(n)-
X*inv(X`*X)*X`)*Y_v_perm[,h])/(n-2-2);
    F_perm[h] = fa[h]/fa[h];
end;

/* Calculate the permuted p value */
p_value_j[j] = count / simul;

```

```

ifp_value_j[j] <0.05then reject = reject + 1;
end;
end;
quit;

```

13.3 Code for Section 9.7

```

title'Simulation of Multiple Linear Regression Procedures';

prociml;
resetnolog;

n=8; /*{9,18,36,54,72,90}*/
nfact = 999;
simul = 5000;

p value = J(simul,1,0);
reject = 0;

/*For each i, xs are generated from a uniform distribution with mean
1 and variance 2,3,5,6
Generate the reference sample*/;
x1=J(n,1,0);
x2=J(n,1,0);
x3=J(n,1,0);
x4=J(n,1,0);
one = J(n,1,1);
e=J(n,1,0);
y=J(n,1,0);
X = {2.5997294    1.5212853    1.3151689    3.4309603,
2.5858015    3.2893621    1.7899367    5.2750796,
2.7751864    1.075762    3.779272    4.9380381,
2.1413    1.6342128    2.6435113    1.4149492,
2.244165    2.0337241    1.2331845    2.1343334,
2.1431852    3.8966126    1.7310676    2.3856989,
1.1584636    2.3290912    1.8921711    6.1513217,
1.1803719    3.7072361    4.1454367    2.6039629
} ;
/*Generate the errors from a normal, exponential, t and uniform
distribution*/

e = {2.6876018,
1.0303678,
1.9167301,
2.3493892,
1.0816412,
1.0312045,
1.6106843,
1.6316914
};
/* X = x1||x2||x3||x4;*/
XH = X[,1:2];
b1 = {1};
b2 = {1};
b3 = {0};
b4 = {0};
b = b1//b2//b3//b4;
y = X*b + e;

```

```

y_perm = J(n,nfact,0);
u = J(n,1,0);
F_perm_a = J(nfact,1,0);
F_perm_b = J(nfact,1,0);
F_perm = J(nfact,1,0);
Fa = J(1,1,0);
Fb = J(1,1,0);
F_ref = J(1,1,0);
M=J(n,n,0);
b0_perm = J(2,nfact,0);
XH = X[,1:2];
X0 = X[,3:4];

do s = 1tosimul;
  /*Generate the y permutations*/
  do j = 1tonfact;
    do l = 1to n;
      u[l] = rannor(0);
    end;

    call SORTNDX( ndx, u, {1}, {1} );
    u = u[ndx,];

    do i = 1to n;
      y_perm[i,j] = y[ndx[i]];
    end;
  end;
  /* print y_perm;*/

  /*Generate the reference sample F*/
  Fa = (y`*(X`inv(X`*X)*X`- XH`inv(XH`*XH)*XH`)*y)/2;
  Fb = (y`*(I(n)- X`inv(X`*X)*X`)*y)/(n-2-2);
  F_ref = Fa/Fb;
  ;

  /* Manly Method */

  /*Calculate the permuted F values*/
  do k = 1tonfact;
    F_perm_a[k] = (y_perm[,k]`*(X`inv(X`*X)*X`-
XH`inv(XH`*XH)*XH`)*y_perm[,k])/2;
    F_perm_b[k] = (y_perm[,k]`*(I(n)-
X`inv(X`*X)*X`)*y_perm[,k])/(n-2-2);
    F_perm[k] = F_perm_a[k]/F_perm_b[k];
  end;

  /*Calculate the permuted p value*/
  count = 0;
  do d = 1tonfact;
    IfF_perm[d] >= 3.6then count = count + 1;
  end;

  title'Manly n=8';
  /* print count nfact y;*/
  p value[s] = count/nfact;
  ifp_value[s] <0.05then reject = reject + 1;

```



```

/* Freedman and Lane Method */

/*Calculate the permuted y values from the reduced model*/
do i = 1tonfact;
    y_perm[,i] = b1*x1 + b2*x2 + e_perm[,i];
end;

/*Calculate the permuted b values*/
do h = 1tonfact;
    b_perm[,h] = inv(X`X)*X`*y_perm[,h];
end;

/*Calculate the permuted y values from the full model*/
do k=1tonfact;
    y_perm_2[,k] = X*b_perm[,k] + e_perm[,k];
end;

/*Calculate the permuted F values*/
do g = 1tonfact;
    F_perm_a[g] = (Y_perm_2[,g] `(X*inv(X`X)*X`-
XH*inv(XH`XH)*XH`) *Y_perm_2[,g])/2;
    F_perm_b[g] = (Y_perm_2[,g] `(I(n)-
X*inv(X`X)*X`) *Y_perm_2[,g])/(n-2-2);
    F_perm[g] = F_perm_a[g]/F_perm_b[g];
end;

/*Calculate the permuted p value*/
count = 0;
do d = 1tonfact;
    IfF_perm[d] >= F_refthen count = count + 1;
end;

/*Calculate the permuted p value*/
count = 0;
do d = 1tonfact;
    IfF_perm[d] >= 3.6then count = count + 1;
end;

title'Freedman & Lane n=8';
/* print count nfact y;*/
p value[s] = count/nfact;
ifp_value[s] <0.05then reject = reject + 1;

/* Kennedy Method */

H = XH*inv(XH`XH)*XH`;
VV = J(n,n,1);
VV = I(n) - H;
y_v = VV`*y;
X0_v = VV`*X0*b0;
e_v = VV`*e;
e_v_perm = J(n,nfact,0);
y_v_perm = J(n,nfact,0);
F_perm_a = J(nfact,1,0);
F_perm_b = J(nfact,1,0);
F_perm = J(nfact,1,0);
Fa = J(1,1,0);
Fb = J(1,1,0);
F_ref = J(1,1,0);

```

```

print X0_v;

/*Generate the permuted y values*/;
do i = 1tonfact;
    y_v_perm[,i] = X0_v + e_perm[,i];
end;
X*inv(X`*X)*X`- XH*inv(XH`*XH)*XH`

    /*Generate the permuted F values*/;
do g = 1tonfact;
    F_perm a[g] =
(Y_v_perm[,g]`*(XX0_v*inv(X0_v`*X0_v)*X0_v`)*Y_v_perm[,g])/2;
    F_perm b[g] = (Y_v_perm[,g]`*(I(n)-
X*inv(X`*X)*X`)*Y_v_perm[,g])/(n-2-2);
    F_perm[g] = F_perm_a[g]/F_perm_b[g];
end;

    /*Calculate the permuted p value*/
count = 0;
do d = 1tonfact;
    IfF_perm[d] >= 6.944272then count = count + 1;
end;

title'Kennedy n=8';
/* print count nfact y;*/
p_value[s] = count/nfact;
ifp_value[s] <3.6then reject = reject + 1;

/* Ter Braak */

    /*Calculate the permuted F values*/
do j = 1tonfact;
    F_perm a[j] = (Y_perm[,j]`*(X*inv(X`*X)*X`-
XH*inv(XH`*XH)*XH`)*Y_perm[,j])/2;
    F_perm b[j] = (Y_perm[,j]`*(I(n)-
X*inv(X`*X)*X`)*Y_perm[,j])/(n-2-2);
    F_perm[j] = F_perm_a[j]/F_perm_b[j];
end;

    /*Calculate the permuted p value*/
count = 0;
do d = 1tonfact;
    IfF_perm[d] >= 3.6then count = count + 1;
end;

title'TerBraak n=8';
/* print count nfact y;*/
p_value[s] = count/nfact;
ifp_value[s] <0.05then reject = reject + 1;

/* Tantanawich Method */

    /*Calculate the matrix M and H*/
M = I(n) - XH*inv(XH`*XH)*XH`;
H = I(n) - X*inv(X`*X)*X`;

```

```

/*Calculate the B2 permutations*/
do l = 1tonfact;
    b0_perm[,l] = b0 + inv(X0`*M*X0)*X0`*M*e_perm[,l];
end;

/*Calculate to F reference statistic*/
Fa = b0`*(X0`*M*X0)*b0;
Fb = y`*H*y;
F_ref = Fa/Fb;

/*Calculate the permuted F statistic*/
do j = 1tonfact;
    F_perm_a[j] = b0_perm[,j]`*(X0`*M*X0)*b0_perm[,j];
    F_perm_b[j] = y`*H*y;
    F_perm[j] = F_perm_a[j]/F_perm_b[j];
end;

/*Calculate the permuted p value*/
count = 0;
do d = 1tonfact;
    IfF perm[d] >= 3.6
then count = count + 1;
end;

title'Tantanawich n=8';
/* print count nfact y;*/
p value[s] = count/nfact;
ifp_value[s] <0.05then reject = reject + 1;

/* Kherad-Pajouh and Renaud */

H = J(n,n,0);
VV = J(n,n,0);
XH = X[,1:2];
X0 = X[,3:4];
H = XH*inv(XH`*XH)*XH`;
VV = I(n) - H;
calleigen (eval,vec,VV);
V = J(n,n,1);
V = vec*(diag(eval)**0.5)*vec`;
y_v = V`*y;
X0_v = V*X0;
e_v = V`*e;

/*Generate the permuted y values*/;
do g = 1tonfact;
    y_v_perm[,g] = X0_v*b0 + e_perm[,g];
end;

/*Generate the permuted F values*/;
do h = 1tonfact;
    F_perm_a[h] =
(Y_v_perm[,h]`*(X0_v*inv(X0_v`*X0_v)*X0_v`)*Y_v_perm[,h])/2;
    F_perm_b[h] = (Y_v_perm[,h]`*(I(n)-
X*inv(X`*X)*X`)*Y_v_perm[,h])/(n-2-2);
    F_perm[h] = F_perm_a[h]/F_perm_b[h];
end;

/*Calculate the permuted p value*/
count = 0;

```

```

do d = 1tonfact;
    IfF_perm[d] >= 3.6
then count = count + 1;
end;

title 'Kherad-Pajouh and Renaud n=8';
/* print count nfact y;*/
p_value[s] = count/nfact;
ifp_value[s] <0.05then reject = reject + 1;
end; /*simulations*/
printp_value reject;
quit

```

13.4 Code for Section 10.5

```

title 'IAAF national track results for the 1984 Los Angeles Olympics';

optionsvalidvarname=any;
odsgraphicson;

data track;
input country $ 1-17 m100 18-24 m200 25-32 m400 33-40 m800 41-48 m1500
49-56 m5000 57-64 m10000 65-73 marathon 74-80;
/* Country 100 m (s) 200 m (s) 400 m (s) 800 m (min) 1,500 m
(min) 5,000 m (min) 10,000 m (min) Marathon (min)*/

datalines;
Argentina0000000010.390020.8100046.84000108.6000222.0000864.00001581.600082
63.20
Australia 10.31 20.06 44.84 104.40 214.20 796.80
1659.60 7698.00
Austria 10.44 20.81 46.82 107.40 216.00 795.60
1663.20 8154.00
Belguim 10.34 20.68 45.04 103.80 216.00 793.20
1647.00 7797.00
Bermuda 10.28 20.58 45.91 108.00 225.00 880.80
1833.00 8797.20
. . . . .
.
. . . . .
.
. . . . .
.
Taipei 10.59 21.29 46.80 107.40 226.20 844.20
1804.20 8356.20
Thailand 10.39 21.09 47.91 109.80 230.40 913.80
1959.00 9594.00
Turkey 10.71 21.43 47.60 107.40 220.20 813.60
1714.80 7890.00
USA 9.93 19.75 43.86 103.80 211.80 792.00
1645.80 7693.20
USSR 10.07 20.00 44.60 105.00 215.40 792.00
1651.80 7833.00
Western Samoa 10.82 21.86 49.00 121.20 257.40 976.80
2082.60 9709.80
;
/* Principal Component Analysis of the Original Data*/

```

```

procprincomp data=Track;
odsselect EigenvaluePlot;
var m100 m200 m400 m800 m1500 m5000 m10000 marathon;
run;
/* Permutation Approach*/

proc iml;
resetnolog;
use track;
read all var {m100, m200, m400, m800, m1500, m5000, m10000, marathon} into
principal;
nfact = 1;
n = 55;
u = J(55, 1, 0);
p_perm = J(55, nfact, 0);
principal_2 = J(55, 8*8+8*8+8*8, 0);
/*print principal;*/
t=0;
do q = 2 to 3;
do k = 1 to 8;
col = J(55, 1, 0);
col = principal[,k];
/*Generate the y permutations*/
do j = 1 to nfact;
do l = 1 to n;
u[l] = rannor(0);
end;

call SORTNDX( ndx, u, {1}, {1} );
u = u[ndx,];

do i = 1 to n;
p_perm[i,j] = col[ndx[i]];
end;
end;
print p_perm;

principal_2[,1:8] = p_perm||principal[,2:8];
principal_2[,8+1:8*2] = principal[,1]||p_perm||principal[,3:8];
principal_2[,8*2+1:8*3] = principal[,1:2]||p_perm||principal[,4:8];
principal_2[,8*3+1:8*4] = principal[,1:3]||p_perm||principal[,5:8];
principal_2[,8*4+1:8*5] = principal[,1:4]||p_perm||principal[,6:8];
principal_2[,8*5+1:8*6] = principal[,1:5]||p_perm||principal[,7:8];
principal_2[,8*6+1:8*7] = principal[,1:6]||p_perm||principal[,8:8];
principal_2[,8*7+1:8*8] = principal[,1:7]||p_perm;
t=q;
/*Q=2*/;
If k = 1 & t = q then principal_2[, (q-1)*8**2+1 : (q-1)*8**2+8*1] =
p_perm||principal[,2:8];
If k = 2 & t = q then principal_2[, (q-1)*8**2+8*1+1: (q-1)*8**2+8*2] =
principal[,1]||p_perm||principal[,3:8];
If k = 3 & t = q then principal_2[, (q-1)*8**2+8*2+1: (q-1)*8**2+8*3] =
principal[,1:2]||p_perm||principal[,4:8];
If k = 4 & t = q then principal_2[, (q-1)*8**2+8*3+1: (q-1)*8**2+8*4] =
principal[,1:3]||p_perm||principal[,5:8];
If k = 5 & t = q then principal_2[, (q-1)*8**2+8*4+1: (q-1)*8**2+8*5] =
principal[,1:4]||p_perm||principal[,6:8];
If k = 6 & t = q then principal_2[, (q-1)*8**2+8*5+1: (q-1)*8**2+8*6] =
principal[,1:5]||p_perm||principal[,7:8];

```

```
If k = 7& t = q then principal_2[(q-1)*8**2+8*6+1:(q-1)*8**2+8*7] =  
principal[,1:6]||p_perm||principal[,8:8];  
If k = 8& t = q then principal_2[(q-1)*8**2+8*7+1:(q-1)*8**2+8**2] =  
principal[,1:7]||p_perm;  
end;  
print principal_2;  
  
create par from principal_2;  
appendfrom principal_2;  
end;  
quit;
```