Chapter 1
Introduction

Abstract This chapter provides an introduction to permutation statistical methods and an overview of the next 10 chapters. The contents of each chapter are described and summarized in considerable detail.

The primary purpose of this book is to introduce the reader to a wide variety of elementary permutation statistical methods. Most readers will be familiar with conventional statistical methods under the Neyman–Pearson population model of statistical inference, such as tests of hypotheses, confidence intervals, simple linear correlation and regression, one-way completely-randomized analysis of variance, one-way randomized-blocks analysis of variance, and chi-squared tests of goodness-of-fit and independence. However, corresponding permutation statistical tests and measures will almost certainly be less familiar to most readers. While permutation methods date back almost 100 years to the early works by R.A. Fisher and E.J.G. Pitman in the 1920s and 1930s, permutation methods are computationally-intensive methods and it took the advent of high-speed computing to make most permutation methods feasible. Thus, permutation statistical methods have emerged as a practical alternative to conventional statistical methods only in the last 30 or so years. Consequently, permutation statistical methods are seldom taught in introductory courses and there exist no introductory-level textbooks on permutation methods at this writing.[1]

Three main themes characterize the 11 chapters of this book. First, test statistic δ is introduced, defined, and detailed. Test statistic δ is the fundamental test statistic for permutation statistical methods and serves both as a replacement for many conventional statistics such as the one-sample t test, the two-sample t test, the matched-pairs t test, the complete range of completely-randomized and randomized-blocks analysis of variance F tests, and a large number of parametric

---

[1]Some introductory textbooks in statistics now include a chapter on permutation methods. For example, an introductory book by Howell titled *Statistical Methods for Psychology* contains a chapter on "Resampling and Nonparametric Approaches to Data" that includes examples of exact and Monte Carlo permutation methods as well as bootstrapping [8].

and nonparametric tests of differences and measures of association and correlation. Moreover, test statistic δ lends itself to the development of new statistical tests and measures. As such, test statistic δ is central to the permutation analyses presented in Chaps. 5–11 and constitutes a unifying test statistic for many permutation-based statistical methods.

Second, measures of effect size have become increasingly important in the reporting of contemporary research with many journals now requiring both tests of significance and associated measures of effect size. Measures of effect size indicate the strength of a statistical difference or relationship. In brief, measures of effect size provide information pertaining to the *practical* or *clinical* significance of a result as contrasted with the *statistical* significance of a result. The two are more often than not reported in concert. Conventional measures of effect size typically belong to one of the two families: the d family or the r family. Measures of effect size in the d family typically report the effect size in standard deviation units with values between 0 and ∞, which is perfectly acceptable when comparing two or more studies but may be difficult to interpret for a single, stand-alone study. Cohen's $\hat{d}$ is probably the best-known measure of effect size in the eponymous d family. Measures of effect size in the r family report the effect size as some variety of squared correlation coefficient with values between 0 and 1. Unfortunately, under many circumstances members of the r family cannot achieve the maximum value of 1. When the maximum value is unknown, it is impossible to interpret intermediate values. Pearson's r^2 coefficient of determination is an example of a measure of effect size in the r family and is the measure from which the family gets its name.

A relatively new measure of effect size based on test statistic δ is introduced and described. Effect size measure $\Re$ is a permutation-based, chance-corrected measure of effect size. Chance-corrected measures have much to commend them as they provide interpretations that are easily understood by the average reader. Positive values indicate an effect size greater than expected by chance, negative values indicate an effect size less than expected by chance, and a value of zero indicates an effect size corresponding to chance. The $\Re$ family of measures of effect sizes serves as a replacement for both the d and r families, including Cohen's $\hat{d}$, Pearson's r^2, Kelley's ϵ^2, and Hays' $\hat{\omega}^2$. As such, effect size measure $\Re$ is central to the permutation analyses presented in Chaps. 5–11 and constitutes a generalized, unifying measure of effect size for many permutation-based statistical methods.

Third, conventional statistics, under the Neyman–Pearson population model of statistical inference, necessarily assume normality. The normal distribution is a two-parameter distribution in which the two parameters are the population mean denoted by μ_x and the population variance denoted by σ_x^2. For most parametric tests the population mean is estimated by the sample mean denoted by $\bar{x}$ and the population variance by the sample variance denoted by s_x^2. The sample mean is the point about which the sum of squared deviations is minimized and the sample variance is the average of the squared deviations about the sample mean. Thus, because of the assumption of normality, squared deviations among sample values are an integral and necessary component of most parametric tests under the Neyman–Pearson population model of statistical inference.

On the other hand, statistical tests and measures under the Fisher–Pitman permutation model are distribution-free, do not assume normality, and because they do not depend on squared deviations among sample values, are not limited to squared deviations about the mean. While any scaling factor can be used with permutation statistical methods, ordinary Euclidean scaling has proven to be the most justifiable. Ordinary Euclidean scaling allows permutation statistical methods to minimize, or completely eliminate, the influence of extreme values or statistical outliers, without having to trim, Winsorize, transform, or convert raw scores to ranks. Moreover, ordinary Euclidean scaling allows geometric consistency between the observation space and the analysis space. Finally, ordinary Euclidean scaling has an intuitive appeal that is absent in squared Euclidean scaling. Analyses in Chaps. 5–11 utilize both squared Euclidean scaling, on which conventional statistics rely, and ordinary Euclidean scaling, when appropriate. The squared and ordinary Euclidean scaling results are then compared and contrasted.

These three constructs, test statistic δ, effect size measure $\Re$, and ordinary Euclidean scaling, constitute the main underpinning structures of the book. Each of the substantive chapters is organized around the three constructs and each construct is compared with conventional test statistics, other measures of effect size, and squared Euclidean scaling, when appropriate.

1.1 Overviews of Chapters 2–11

This chapter provides an overview of the book and brief summaries of the following 10 chapters. The format of the book follows the conventional structure of most introductory textbooks in statistical methods with chapters on central tendency and variability, one- and two-sample tests, multi-sample tests, linear correlation and regression, and the analysis of contingency tables. No statistical background of the reader is assumed other than an introductory course in basic statistics, such as is taught in departments of statistics, mathematics, business, biology, economics, or psychology. No mathematical expertise of the reader is assumed beyond elementary algebra.

Most of the substantive chapters in this book follow the same format wherein six example analyses based on permutation statistical methods are provided. The first example in each chapter introduces the main permutation test statistic for the chapter and provides both a highly detailed exact permutation analysis and a conventional analysis; for example, a one-sample permutation test of the null hypothesis under the Fisher–Pitman model and Student's conventional one-sample t test of the null hypothesis under the Neyman–Pearson model. The second example introduces appropriate conventional measures of effect size, for example, Cohen's $\hat{d}$ or Pearson's r^2, and provides a permutation-based, chance-corrected alternative measure of effect size. Because conventional statistical methods under the Neyman–Pearson population model assume random sampling from a normally distributed population, squared deviations about the mean are necessary. Statistical methods

under the Fisher–Pitman permutation model do not assume normality; thus, the third example compares permutation analyses based on ordinary and squared Euclidean scaling functions. The inclusion of one or more extreme values demonstrates the advantages of ordinary Euclidean scaling.

The fourth example introduces Monte Carlo permutation statistical methods wherein a large random sample of all possible permutations is generated and analyzed, in contrast to exact permutation methods wherein all possible permutations are generated and analyzed. Both exact and Monte Carlo permutation analyses are compared with each other and with a conventional statistical analysis. The fifth example applies permutation statistical methods to rank-score data, comparing a permutation statistical analysis to a conventional statistical analysis; for example, a permutation test for two sets of rank scores and the Wilcoxon–Mann–Whitney rank-sum test. The sixth example applies permutation statistical methods to multivariate data, comparing a permutation statistical analysis with a conventional statistical analysis; for example, a permutation test of multivariate matched pairs and Hotelling's multivariate T^2 test for two matched samples.

1.2   Chapter 2

The second chapter provides a brief history of the origins and subsequent development of permutation statistical methods. Permutation statistical methods are a paradox of old and new. While permutation methods predate many conventional parametric statistical methods, only recently have permutation methods become part of the mainstream discussion regarding statistical testing. Permutation statistical methods were introduced by R.A. Fisher in 1925 by calculating an exact probability value using the binomial probability distribution [4]. In 1927 R.C. Geary used an exact permutation analysis to demonstrate the utility of asymptotic approaches for data analysis in an investigation of the properties of linear correlation and regression in finite populations [6].

In 1933 T. Eden and F. Yates examined height measurements of wheat shoots grown in eight blocks. Simulated and theoretical probabilities based on the normality assumption were compared and found to be in close agreement, supporting the assumption of normality [3]. In 1936 H. Hotelling and M.R. Pabst used permutation statistical methods to calculate exact probability values for small samples of ranked data in an examination of correlation methods [7]. In 1937 and 1938 E.J.G. Pitman contributed three seminal papers on permutation statistical methods. The first paper utilized permutation statistical methods in an analysis of two independent samples, the second paper utilized permutation statistical methods in an analysis of linear correlation, and the third paper utilized permutation statistical methods in an analysis of randomized-blocks analysis of variance designs [14–16].

The 1940s and 1950s witnessed a proliferation of nonparametric rank tests. For example, Wilcoxon's two-sample rank-sum test in 1945 [17], Mann and Whitney's two-sample rank-sum test in 1947 [11], Kendall's book on *Rank Correlation*

Methods in 1948 [9], Freeman and Halton's exact methods for analyzing two-way and three-way contingency tables in 1951 [5], Kruskal and Wallis' *C*-sample rank-sum test in 1952 [10], Box and Andersen's promotion of permutation methods in the derivation of robust criteria in 1955 [1], and Dwass's rigorous investigation into the precision of Monte Carlo permutation methods in 1957 [2]. In many of these papers, permutation methods were employed to generate tables of exact probability values for small samples.

In the 1960s and 1970s mainframe computers became available to researchers at major universities and by the end of the period desktop computers and workstations, although not common, were available to many investigators. In addition, the speed of computing increased greatly between 1970 and 1980. Permutation statistical methods arrived at a level of maturity during the period 1980–2000 primarily as a result of two factors: greatly improved computer clock speeds and widely-available desktop computers and workstations. By the early 2000s, computing power had advanced enough that permutation statistical methods were providing exact probability values in an efficient manner for a wide variety of statistical tests and measures [12, 13].

1.3 Chapter 3

The third chapter opens with a description of two models of statistical inference: the well-known and widely-taught Neyman–Pearson population model and the lesser-known and seldom-taught Fisher–Pitman permutation model. Under the permutation model, three types of permutation methods are described: exact permutation methods yielding precise probability values, Monte Carlo permutation methods yielding approximate but highly accurate probability values, and per-mutation methods based on moment approximations yielding exact moments and approximate probability values. In this chapter the Neyman–Pearson population model and Fisher–Pitman permutation model are compared and contrasted and the advantages of permutation statistical methods are described.

Because permutation methods are computationally intensive methods, often requiring millions of calculations, five computational efficiencies are described in Chap. 3. First, high-speed computing and, in the case of Monte Carlo permutation methods, efficient pseudo-random number generators. Second, the examination of all combinations instead of all permutations of the observed data. Third, the use of mathematical recursion. Fourth, calculation of only the variable portion of the selected test statistic. Fifth, in the case of multiple arrays of data, holding one array of the observed data constant. Where appropriate, each efficiency is described and illustrated with a small set of data and an example permutation analysis.

1.4 Chapter 4

The fourth chapter provides a general introduction to measures of central tendency and variability, two concepts that are central to conventional statistical analysis and inference. The sample mode, mean, and median are described and illustrated with small example data sets. The sample mode is simply the score or category with the largest frequency. Two example analyses illustrate the mode, one employing scores and the other employing categories.

Next, the sample mean is considered. The sample mean is the point about which the sum of deviations is zero and, more importantly, the point about which the sum of squared deviations is minimized. These properties are illustrated with two example analyses. Moreover, the sample mean is central to the sample standard deviation, denoted by s_x, and the sample variance, denoted by s_x^2— a point that is illustrated with a small set of example data.

The sample median is usually defined as the point below which half the ordered values fall or the 50th percentile. More importantly, the median is the point about which the sum of absolute deviations is minimized. A detailed example analysis illustrates this property. The sample median is central to the mean absolute deviation (MAD), which is illustrated with a small set of example data.

Finally, the mean, median, and mode are compared with each other and an alternative approach to the mean and median based on paired differences is presented and illustrated. The paired-differences approach to the mean and median is central to the Fisher–Pitman permutation model of statistical inference.

1.5 Chapter 5

The fifth chapter provides a general introduction to permutation analyses of one-sample tests of hypotheses. One-sample tests are the simplest of a large family of tests. For this reason, Chap. 5 is the first chapter dealing with the more technical aspects of permutation statistical methods, serves as an introduction to the basic concepts and varieties of permutation statistical methods, and lays a conceptual foundation for subsequent chapters.

First, Chap. 5 defines permutation test statistic δ for one-sample tests, establishes the relationship between test statistic δ and Student's conventional one-sample t test statistic, and describes the permutation procedures for determining exact probability values under the Fisher–Pitman null hypothesis. An example analysis with a small set of data details the required calculations for an exact test of the null hypothesis under the Fisher–Pitman permutation model of statistical inference.

Second, Chap. 5 introduces the concept of effect sizes: indices to the magnitudes of treatment effects and the practical—in contrast to the statistical—significance of the research. The development and publication of measures of effect size has become increasingly important in recent years and a number of journals now require

measures of effect size prior to publication. Three types of measures of effect size are described in Chap. 5. The first type of measure of effect size, designated the d family, is based on measurements of the differences among treatment groups or levels of an independent variable. As noted previously, Cohen's $\hat{d}$ is the most prominent member of the d family, which typically measures effect size by the number of standard deviations separating the means of treatment groups. Thus Cohen's $\hat{d}$ can potentially vary from 0 to ∞.

The second type of measure of effect size, designated the r family, represents some sort of relationship among variables. Measures of effect size in the r family are typically measures of correlation or association, the most familiar being Pearson's squared product-moment correlation coefficient, denoted by r^2. The principle advantage of r measures of effect size is that they are usually bounded by the probability limits 0 and 1, making them easily interpretable.

The third type of measure of effect size, designated the $\Re$ family, represents chance-corrected measures of effect size. Chance-corrected measures are easily understood by the average reader, where positive values indicate an effect size greater than expected by chance, negative values indicate an effect size less than expected by chance, and a value of zero indicates an effect size corresponding to chance. The interrelationships among Student's one-sample t test, Cohen's $\hat{d}$ measure of effect size, Pearson's r^2 measure of effect size, and Mielke and Berry's $\Re$ chance-corrected measure of effect size are explored and illustrated with a small example set of data.

Third, six illustrative examples are provided in Chap. 5, demonstrating permutation statistical methods for one-sample tests of hypotheses. The first example utilizes a small set of data to describe the calculations required for test statistic δ and an exact permutation analysis of a one-sample test under the Fisher–Pitman null hypothesis. Permutation test statistic δ is developed for the analysis of a single sample and compared with Student's conventional one-sample t test.

The second example details measures of effect size for one-sample tests. Specifically, Cohen's $\hat{d}$ and Pearson's r^2 measures of effect size are detailed and $\Re$, an alternative permutation-based, chance-corrected measure of effect size is described for one-sample tests. The differences among the three measures of effect size and their interrelationships are explored and illustrated with a small set of data.

The third example is designed to illustrate the differences between permutation analyses based on ordinary and squared Euclidean scaling functions. Unlike conventional statistical tests that assume normality and are therefore limited to squared Euclidean scaling functions, permutation statistical tests do not assume normality, are extremely flexible, and can accommodate a variety of scaling functions. Inclusion of extreme values illustrates the impact of extreme values on the two scaling functions, on Student's t test statistic, on test statistic δ, on the $\Re$ measure of effect size, and on exact and asymptotic probability values.

The fourth example compares and contrasts exact and Monte Carlo permutation statistical methods. When sample sizes are large, exact permutation tests become impractical and Monte Carlo permutation tests become necessary. While exact permutation tests examine all possible arrangements of the observed data, Monte

Carlo permutation tests examine only a random sample of all possible arrangements of the observed data. Monte Carlo sample sizes can be increased to yield probability values to any desired accuracy, at the expense of computation time.

The fifth example illustrates permutation statistical methods applied to univariate rank-score data. The conventional one-sample tests for rank-score data under the Neyman–Pearson population model are Wilcoxon's signed-rank test and the simple sign test. Wilcoxon's signed-rank test and the sign test are described and compared with permutation-based alternatives. The permutation analyses incorporate both ordinary and squared Euclidean scaling functions. Test statistic δ is defined for rank-score data, the exact probability of δ is generated, and the $\Re$ measure of effect size is described for univariate rank-score data.

The sixth example illustrates permutation statistical methods applied to multivariate data. Multivariate tests have become very popular in recent years as they preserve the relationship among variables, instead of combining the variables into an index and then employing a univariate one-sample test. Like the previous examples, the multivariate permutation analysis incorporates both ordinary and squared Euclidean scaling functions. Test statistic δ is defined for multivariate data, the exact probability of δ is generated, and the $\Re$ measure of effect size is described for multivariate one-sample tests.

1.6   Chapter 6

The sixth chapter provides a general introduction to two-sample tests of hypotheses. Tests of experimental differences for two independent samples are ubiquitous in the research literature and are the tests of choice for comparing control and treatment groups in experimental designs and for comparing two unrelated groups of subjects in survey research.

First, Chap. 6 defines permutation test statistic δ for two independent samples, establishes the relationship between test statistic δ and Student's conventional t test statistic for two independent samples, and describes the permutation procedures for determining exact probability values under the Fisher–Pitman null hypothesis. A small example analysis details the calculations required for an exact test of the null hypothesis under the Fisher–Pitman permutation model of statistical inference.

Second, Chap. 6 describes five measures of effect size for two independent samples. Specifically, Cohen's $\hat{d}$, Pearson's r^2, Kelley's ϵ^2, Hays' $\hat{\omega}^2$, and Mielke and Berry's $\Re$ measures of effect size are described and the interrelationships among t, $\hat{d}$, r^2, ϵ^2, $\hat{\omega}^2$, and $\Re$ are explored and illustrated with a small set of data.

Third, six illustrative examples are provided in Chap. 6, demonstrating permutation statistical methods for tests of two independent samples. The first example utilizes a small data set to detail the calculations required for test statistic δ and an exact permutation test for two independent samples under the Fisher–Pitman null hypothesis. Permutation test statistic δ is developed for the analysis of two independent samples and compared with Student's conventional two-sample t test.

The second example illustrates measures of effect size for two-sample tests. Four conventional measures of effect size are described: Cohen's $\hat{d}$, Pearson's r^2, Kelley's ϵ^2, and Hays' $\hat{\omega}^2$. The four measures are compared and contrasted with Mielke and Berry's $\Re$ chance-corrected measure of effect size.

The third example illustrates the differences between permutation analyses based on ordinary and squared Euclidean scaling functions. The inclusion of extreme values illustrates the impact of extreme values on the two scaling functions, on Student's t test statistic for two independent samples, on test statistic δ, on the $\Re$ measure of effect size, and on exact and asymptotic probability values.

The fourth example compares and contrasts exact and Monte Carlo permutation statistical methods for tests of two independent samples. Both ordinary and squared Euclidean scaling functions are included and evaluated. Finally, the chance-corrected effect size measure $\Re$ is compared with Cohen's $\hat{d}$, Pearson's r^2, Kelley's ϵ^2, and Hays' $\hat{\omega}^2$ measures of effect size.

The fifth example illustrates permutation statistical methods applied to univariate rank-score data. The conventional two-sample test for rank scores under the Neyman–Pearson population model is the Wilcoxon–Mann–Whitney (WMW) two-sample rank-sum test. The WMW test is described and compared with alternative tests under the Fisher–Pitman permutation model. The permutation analyses incorporate both ordinary and squared Euclidean scaling functions. Test statistic δ is defined for rank-score data, the exact and Monte Carlo probability values for δ are developed, and the $\Re$ measure of effect size is described for univariate rank-score data.

The sixth example illustrates permutation statistical methods applied to multivariate data. The results of a permutation statistical analysis are compared with the results from Hotelling's multivariate T^2 test for two independent samples. Mielke and Berry's $\Re$ chance-corrected measure of effect size is described and illustrated for multivariate data.

1.7   Chapter 7

The seventh chapter provides a general introduction to matched-pairs tests of hypotheses. Tests of experimental differences between two matched samples are the simplest of a very large family of tests. In general, matched-pairs tests generally possess more power than tests for two independent samples with the same number of subjects or the same power with fewer subjects. In addition, matched-pairs tests are always balanced with the same number of subjects in each treatment group, a decided advantage over conventional tests for two independent samples, where the two samples may be markedly different in size.

First, Chap. 7 introduces permutation test statistic δ for matched-pairs tests, establishes the relationship between test statistic δ and Student's matched-pairs t test statistic, and describes the permutation procedures required for determining exact probability values under the Fisher–Pitman null hypothesis. Permutation test

statistic, δ, is developed for the analysis of two matched samples and compared with Student's conventional matched-pairs t test.

Second, Chap. 7 describes measures of effect size for matched-pairs tests. Specifically, Student's t test for matched pairs, Cohen's $\hat{d}$, Pearson's r^2, and Mielke and Berry's $\Re$ measure of effect size are presented and the interrelationships among t, $\hat{d}$, r^2, and $\Re$ are explored and illustrated with a small example set of data.

Third, six illustrative examples are provided in Chap. 7, demonstrating permutation statistical methods for matched-pairs tests. The first example utilizes a small set of data to detail the calculations required for test statistic δ and an exact permutation test for matched pairs under the Fisher–Pitman null hypothesis. Permutation test statistic δ is developed for the analysis of matched pairs and compared with Student's conventional matched-pairs t test statistic.

The second example describes measures of effect size for matched-pairs tests. Cohen's $\hat{d}$ and Pearson's r^2 measures of effect size are described and Mielke and Berry's chance-corrected measure of effect size, $\Re$, is developed for matched-pairs analyses and compared with Cohen's $\hat{d}$ and Pearson's r^2 conventional measures of effect size.

The third example illustrates the differences between analyses based on ordinary and squared Euclidean scaling functions. Inclusion of extreme values underscores the impact of extreme values on the two scaling functions, on Student's t test statistic for two matched samples, on test statistic δ, on the $\Re$ measure of effect size, and on the accuracy of exact and asymptotic probability values.

The fourth example compares and contrasts exact and Monte Carlo permutation analyses for matched-pairs tests. A matched-pairs test with a large data set is utilized to generate exact and Monte Carlo permutation tests for both ordinary and squared Euclidean scaling functions. The example confirms that Monte Carlo permutation tests are a suitable and efficient substitute for exact permutation tests, provided the Monte Carlo random sample arrangement of the observed data is sufficiently large. Finally, the $\Re$ measure of effect size is described for matched-pairs tests and compared with Cohen's $\hat{d}$ and Pearson's r^2 conventional measures of effect size.

The fifth example illustrates permutation statistical methods applied to univariate rank-score data, comparing permutation statistical methods to Wilcoxon's conventional signed-ranks test and the sign test. A large matched-pairs data set is utilized to generate both exact and Monte Carlo permutation tests for both ordinary and squared Euclidean scaling functions. Finally, the $\Re$ measure of effect size is described and illustrated for univariate rank-score data.

The sixth example illustrates permutation statistical methods applied to multivariate matched-pairs data. Test statistic δ is shown to be related to Hotelling's conventional T^2 test for matched pairs with a squared Euclidean scaling function. The results for test statistics δ and T^2 are compared. Finally, Mielke and Berry's $\Re$ measure of effect size is described and illustrated for multivariate data.

1.8   Chapter 8

The eighth chapter presents permutation statistical methods for analyzing experimental differences among three or more independent samples, commonly called completely-randomized designs under the Neyman–Pearson population model. Multi-sample tests are of two types: tests for differences among three or more independent samples (completely-randomized designs) and tests for differences among three or more related samples (randomized-blocks designs). Permutation statistical tests for multiple independent samples are described in Chap. 8 and permutation statistical tests for multiple related samples are described in Chap. 9.

Six example analyses illustrate permutation statistical methods for multi-sample tests. The first example utilizes a small set of data to illustrate the calculations required for test statistic δ and an exact permutation test for multiple independent samples under the Fisher–Pitman null hypothesis. Permutation test statistic, δ, is developed for the analysis of multiple independent samples and compared with Fisher's conventional F-ratio test statistic for completely-randomized designs.

The second example develops the $\Re$ measure of effect size as a chance-corrected alternative to the four conventional measures of effect size for multi-sample tests: Cohen's $\hat{d}$, Pearson's η^2, Kelley's $\hat{\eta}^2$, and Hays' $\hat{\omega}^2$.

The third example compares permutation statistical methods based on ordinary Euclidean scaling functions with permutation methods based on squared Euclidean scaling functions. Inclusion of one or more extreme scores underscores the impact of extreme values on the two scaling functions, on Fisher's F-ratio test statistic for completely-randomized designs, on the permutation test statistic δ, on the $\Re$ measure of effect size, and on the accuracy of exact and asymptotic probability values.

The fourth example compares and contrasts exact and Monte Carlo permutation methods for multiple independent samples. Both ordinary and squared Euclidean scaling functions are evaluated. Finally, the $\Re$ measure of effect size is compared with the four conventional effect size measures for multi-sample tests: Cohen's $\hat{d}$, Pearson's η^2, Kelley's $\hat{\eta}^2$, and Hays' $\hat{\omega}^2$.

The fifth example illustrates the application of permutation statistical methods to univariate rank-score data, comparing a permutation analysis of example data to the conventional Kruskal–Wallis one-way analysis of variance for ranks test. Both exact and Monte Carlo permutation analyses are utilized and compared. Mielke and Berry's chance-corrected $\Re$ measure of effect size is described and illustrated for univariate rank-score data.

The sixth example illustrates the application of permutation statistical methods to multivariate data, comparing a permutation analysis of example data to the conventional Bartlett–Nanda–Pillai trace test for multivariate data. Mielke and Berry's chance-corrected $\Re$ measure of effect size is described for multivariate data and compared with η^2, the conventional measure of effect size for multivariate data.

1.9   Chapter 9

The ninth chapter presents permutation statistical methods for analyzing experimental differences among three or more matched samples, commonly called randomized-blocks designs under the Neyman–Pearson population model. Randomized-blocks constitute important research designs in many fields. In recent years randomized-blocks designs have become increasingly important in fields such as horticulture, animal science, and agronomy as it has become easier to produce matched subjects through embryo transplants, cloning, genetic engineering, and selective breeding.

Six example analyses illustrate the application of permutation statistical methods to randomized-blocks designs. The first example utilizes a small set of data to detail the calculations required for test statistic δ and an exact permutation test for multiple matched samples under the Fisher–Pitman null hypothesis. Permutation test statistic, δ, is developed for the analysis of multiple matched samples and compared with Fisher's conventional F-ratio test for randomized-blocks designs.

The second example develops the $\Re$ measure of effect size as a chance-corrected alternative to the four conventional measures of effect size for randomized-blocks designs: Hays' $\hat{\omega}^2$, Pearson's η^2, Cohen's partial η^2, and Cohen's f^2.

The third example compares permutation statistical methods based on ordinary and squared Euclidean scaling functions. Inclusion of one or more extreme scores underscores the impact of extreme values on the two scaling functions, on Fisher's F-ratio test statistic for randomized-blocks designs, on the permutation test statistic δ, on the $\Re$ measure of effect size, and on the accuracy of exact and asymptotic probability values. It is demonstrated that extreme blocks of data yield the same results with both scaling functions, but extreme values within a block can yield considerable differences.

The fourth example utilizes a larger data set to compare and contrast exact and Monte Carlo permutation statistical methods for randomized-blocks designs. Both ordinary and squared Euclidean scaling functions are evaluated. The chance-corrected measure of effect size $\Re$ is developed for randomized-blocks designs and compared with Hays' $\hat{\omega}^2$, Pearson's η^2, Cohen's partial η^2, and Cohen's f^2 conventional measures of effect size.

The fifth example illustrates the application of permutation statistical methods to univariate rank-score data, comparing permutation statistical methods to Friedman's conventional two-way analysis of variance for ranks. The permutation test statistic, δ, and Mielke and Berry's $\Re$ measure of effect size are described and illustrated for univariate rank-score data.

The sixth example illustrates the application of permutation statistical methods to multivariate randomized-blocks designs. Both the permutation test statistic δ and Mielke and Berry's chance-corrected $\Re$ measure of effect size are described and illustrated for multivariate randomized-blocks designs.

1.10   Chapter 10

The tenth chapter presents permutation statistical methods for measures of linear correlation and regression. Measures of linear correlation and regression are ubiquitous in the research literature and constitute the backbone of many more advanced statistical methods, such as factor analysis, principal components analysis, path analysis, network analysis, neural network analysis, multi-level (hierarchical) modeling, and structural equation modeling.

Six example analyses illustrate the application of permutation statistical methods to linear correlation and regression. The first example utilizes a small set of bivariate observations to illustrate the calculations required for test statistic δ and an exact permutation test for measures of linear correlation under the Fisher–Pitman null hypothesis. Permutation test statistic, δ, is developed for the analysis of correlation and compared with Pearson's conventional squared product-moment correlation coefficient.

The second example develops the $\Re$ measure of effect size as a chance-corrected alternative to Pearson's squared product-moment correlation coefficient. The two measures of effect size are illustrated and compared using a small set of data.

The third example compares permutation statistical methods based on ordinary and squared Euclidean scaling functions, with an emphasis on the analysis of data containing one or more extreme values. Ordinary least squares (OLS) regression based on squared Euclidean scaling and least absolute deviation (LAD) regression based on ordinary Euclidean scaling are described and compared.

The fourth example compares exact and Monte Carlo permutation statistical methods for linear correlation and regression. Both ordinary Euclidean scaling and squared Euclidean scaling functions are evaluated. The chance-corrected effect-size measure $\Re$ is developed for correlation methods and compared with Pearson's squared product-moment correlation coefficient.

The fifth example illustrates the application of permutation statistical methods to univariate rank-score data, comparing permutation statistical methods with Spearman's rank-order correlation coefficient, Kendall's rank-order correlation coefficient, and Spearman's footrule correlation coefficient. The permutation test statistic δ and Mielke and Berry's $\Re$ measure of effect size are described and illustrated for univariate rank-score data.

The sixth example illustrates the application of permutation statistical methods to multivariate linear correlation and regression. Both OLS and LAD multivariate linear regression are described and compared for multivariate observations. Permutation test statistic δ and the $\Re$ measure of effect size are described and illustrated for multivariate linear regression data.

1.11   Chapter 11

The last chapter provides a general introduction to permutation measures of association for contingency tables. Measures of association for contingency tables constitute a variety of types. One type measures the association in a cross-classification of two nominal-level (categorical) variables and the measure can be either symmetric or asymmetric. A second type measures the association in a cross-classification of two ordinal-level (ranked) variables and the measure can be either symmetric or asymmetric. A third type measures the association in a cross-classification of a nominal-level variable and an ordinal-level variable. A fourth type measures the association in a cross-classification of a nominal-level variable and an interval-level variable. And a fifth type measures the association in a cross-classification of an ordinal-level variable and an interval-level variable. These mixed-level measures are typically asymmetric with the lower-level variable serving as the independent variable and the higher-level variable serving as the dependent variable.

Six sections of Chap. 11 illustrate permutation statistical methods for the analysis of contingency tables. The first section considers permutation statistical methods applied to conventional goodness-of-fit tests; for example, Pearson's chi-squared goodness-of-fit test. Two examples illustrate permutation goodness-of-fit tests and a new maximum-corrected measure of effect size is developed for chi-squared goodness-of-fit tests.

The second section considers permutation statistical methods for analyzing contingency tables in which two nominal-level variables have been cross-classified. Cramér's V test statistic illustrates a conventional symmetrical measure of nominal association and Goodman and Kruskal's t_a and t_b illustrate conventional asymmetrical measures of nominal association.

The third section utilizes permutation statistical methods for analyzing contingency tables in which two ordinal-level variables have been cross-classified. Goodman and Kruskal's G test statistic illustrates a conventional symmetrical measure of ordinal association and Somers' d_{yx} and d_{xy} test statistics illustrate conventional asymmetrical measures of ordinal association.

The fourth section utilizes permutation statistical methods for analyzing contingency tables in which one nominal-level variable and one ordinal-level variable have been cross-classified. Freeman's θ test statistic illustrates a conventional measure of nominal-ordinal association.

The fifth section utilizes permutation statistical methods for analyzing contingency tables in which one nominal-level variable and one interval-level variable have been cross-classified. Pearson's point-biserial correlation coefficient illustrates a conventional measure of nominal-interval association.

The sixth section utilizes permutation statistical methods for analyzing contingency tables in which one ordinal-level variable and one interval-level variable have been cross-classified. Jaspen's $r_{Y\bar{Z}}$ correlation coefficient illustrates a conventional measure of ordinal-interval association.

1.12   Summary

This chapter provided an introduction to permutation statistical methods and an overview and brief summaries of the next 10 chapters. Most of the substantive chapters utilize six examples or sections to illustrate the application of permutation statistical methods to one-sample tests, tests for two independent samples, matched-pairs tests, completely-randomized designs, randomized-blocks designs, linear correlation and regression, and a variety of types of contingency tables.

Chapter 2 provides a brief history and subsequent development of permutation statistical methods. Permutation statistical methods were introduced by R.A. Fisher in 1925, further developed by R.C. Geary in 1927, T. Eden and F. Yates in 1933, and H. Hotelling and M.R. Pabst in 1936, but it was E.J.G. Pitman who made permutation statistical methods explicit with three seminal articles published in 1937 and 1938. However, it took another 50 years before high-speed computing was developed and permutation statistical methods became practical.

References

1. Box, G.E.P., Andersen, S.L.: Permutation theory in the derivation of robust criteria and the study of departures from assumption (with discussion). J. R. Stat. Soc. B Meth. **17**, 1–34 (1955)
2. Dwass, M.: Modified randomization tests for nonparametric hypotheses. Ann. Math. Stat. **28**, 181–187 (1957)
3. 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. **23**, 6–17 (1933)
4. Fisher, R.A.: The arrangement of field experiments. J. Am. Stat. Assoc. **33**, 503–513 (1926)
5. Freeman, G.H., Halton, J.H.: Note on an exact treatment of contingency, goodness of fit and other problems of significance. Biometrika **38**, 141–149 (1951)
6. Geary, R.C.: Some properties of correlation and regression in a limited universe. Metron **7**, 83–119 (1927)
7. Hotelling, H., Pabst, M.R.: Rank correlation and tests of significance involving no assumption of normality. Ann. Math. Stat. **7**, 29–43 (1936)
8. Howell, D.C.: Statistical Methods for Psychology, 8th edn. Wadsworth, Belmont (2013)
9. Kendall, M.G.: Rank Correlation Methods. Griffin, London (1948)
10. Kruskal, W.H., Wallis, W.A.: Use of ranks in one-criterion variance analysis. J. Am. Stat. Assoc. **47**, 583–621 (1952). [Erratum: J. Am. Stat. Assoc. **48**, 907–911 (1953)]
11. Mann, H.B., Whitney, D.R.: On a test of whether one of two random variables is stochastically larger than the other. Ann. Math. Stat. **18**, 50–60 (1947)
12. Mielke, P.W., Berry, K.J.: Permutation Methods: A Distance Function Approach. Springer, New York (2001)
13. Mielke, P.W., Berry, K.J.: Permutation Methods: A Distance Function Approach, 2nd edn. Springer, New York (2007)
14. Pitman, E.J.G.: Significance tests which may be applied to samples from any populations. Suppl. J. R. Stat. Soc. **4**, 119–130 (1937)
15. 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. **4**, 225–232 (1937)
16. Pitman, E.J.G.: Significance tests which may be applied to samples from any populations: III. The analysis of variance test. Biometrika **29**, 322–335 (1938)
17. Wilcoxon, F.: Individual comparisons by ranking methods. Biometrics Bull. **1**, 80–83 (1945)