

Chapter 8

Completely-Randomized Designs



Abstract This chapter introduces permutation methods for multiple independent variables; that is, completely-randomized designs. Included in this chapter are six example analyses illustrating computation of exact permutation probability values for multi-sample tests, calculation of measures of effect size for multi-sample tests, the effect of extreme values on conventional and permutation multi-sample tests, exact and Monte Carlo permutation procedures for multi-sample tests, application of permutation methods to multi-sample rank-score data, and analysis of multi-sample multivariate data. Included in this chapter are permutation versions of Fisher's F test for one-way, completely-randomized analysis of variance, the Kruskal–Wallis one-way analysis of variance for ranks, the Bartlett–Nanda–Pillai trace test for multivariate analysis of variance, and a permutation-based alternative for 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' $\hat{\omega}^2$.

This chapter presents exact and Monte Carlo permutation statistical methods for multi-sample tests. Multi-sample tests are of two types: tests for experimental differences among three or more independent samples (completely-randomized designs) and tests for experimental differences among three or more dependent samples (randomized-blocks designs).¹ Permutation statistical methods for multiple dependent samples are presented in Chap. 9. Permutation statistical methods for multiple independent samples are presented in this chapter. In addition there are mixed models with one or more independent samples and one or more dependent samples, but these models are beyond the scope of this introductory book on permutation statistical methods. Interested readers can consult a 2016 book on *Permutation Statistical Methods: An Integrated Approach* by the authors [2].

Multi-sample tests for independent samples constitute a large family of tests in conventional statistical methods. Included in this family are one-way analysis of variance with univariate responses (ANOVA), one-way analysis of variance with

¹In some disciplines tests on multiple independent samples are known as between-subjects tests and tests for multiple dependent or related samples are known as within-subjects tests.

multivariate responses (MANOVA), one-way analysis of variance with one or more covariates and univariate responses (ANCOVA), one-way analysis of variance with one or more covariates and multivariate responses (MANCOVA), and a variety of factorial designs that may be two-way, three-way, four-way, nested, balanced, unbalanced, fixed, random, or mixed.

In this chapter, permutation statistical methods for multiple independent samples are illustrated with six example analyses. The first example utilizes a small set of data to illustrate the computation of exact permutation methods for multiple independent samples, wherein the permutation test statistic, δ , is developed and compared with Fisher's conventional F -ratio test statistic. The second example develops a permutation-based measure of effect size as a chance-corrected alternative to the five conventional measures of effect size for multi-sample tests: Cohen's \hat{d} , Pearson's η^2 , Kelley's $\hat{\eta}^2$, Hays' $\hat{\omega}_F^2$ for fixed models, and Hays' $\hat{\omega}_R^2$ for random models. The third example compares permutation statistical methods based on ordinary and squared Euclidean scaling functions, with an emphasis on the analysis of data sets containing extreme values. The fourth example utilizes a larger data set to provide a comparison of exact permutation methods and Monte Carlo permutation methods, demonstrating the efficiency and accuracy of Monte Carlo statistical methods for multi-sample tests. The fifth example illustrates the application of permutation statistical methods to univariate rank-score data, comparing permutation statistical methods to the conventional Kruskal–Wallis one-way analysis of variance for ranks test. The sixth example illustrates the application of permutation statistical methods to multivariate data, comparing permutation statistical methods with the conventional Bartlett–Nanda–Pillai trace test for multivariate data.

8.1 Introduction

The most popular univariate test for $g \geq 3$ independent samples under the Neyman–Pearson population model of statistical inference is Fisher's one-way analysis of variance wherein the null hypothesis (H_0) posits no mean differences among the g populations from which the samples are presumed to have been randomly drawn; that is, $H_0: \mu_1 = \mu_2 = \dots = \mu_g$. It should be noted that Fisher, writing in the first edition of *Statistical Methods for Research Workers* in 1925, named the aforementioned statistic the variance-ratio test, symbolized it as z , and defined it as

$$z = \frac{1}{2} \log_e \left(\frac{v_1}{v_0} \right),$$

where $v_1 = MS_{\text{Between}}$ and $v_0 = MS_{\text{Within}}$ in modern notation. In 1934, in an effort to eliminate the calculation of the natural logarithm required for calculating Fisher's z test, George Snedecor at Iowa State University published tabled values in a small monograph for Fisher's variance-ratio z statistic and renamed the test statistic F ,

presumably in honor of Fisher [22]. It has often been reported that Fisher was displeased when the variance-ratio z test statistic was renamed F by Snedecor [4, 8].

Fisher's F -ratio test for a completely-randomized design does not determine whether or not the null hypothesis is true, but only provides the probability that, if the null hypothesis is true, the samples have been drawn from populations with identical mean values, assuming normality and homogeneity of variance.

Consider a conventional multi-sample F test with samples of independent and identically distributed univariate random variables of sizes n_1, \dots, n_g , viz.,

$$\{x_{11}, \dots, x_{n_1 1}\}, \dots, \{x_{1g}, \dots, x_{n_g g}\},$$

drawn from g specified populations with cumulative distribution functions $F_1(x), \dots, F_g(x)$, respectively. For simplicity, suppose that population i is normal with mean μ_i and variance σ^2 for $i = 1, \dots, g$. This is the standard one-way classification model with g treatment groups. Under the Neyman–Pearson population model of statistical inference, the null hypothesis of no differences among the population means tests

$$H_0: \mu_1 = \mu_2 = \dots = \mu_g \quad \text{versus} \quad H_1: \mu_i \neq \mu_j \quad \text{for some } i \neq j$$

for g treatment groups. The permissible probability of a type I error is denoted by α and if the observed value of Fisher's F -ratio test statistic is equal to or greater than the critical value of F that defines α , the null hypothesis is rejected with a probability of type I error equal to or less than α , under the assumptions of normality and homogeneity.

For multi-sample tests with g treatment groups and N observations, Fisher's F -ratio test statistic is given by

$$F = \frac{MS_{\text{Between}}}{MS_{\text{Within}}},$$

where the mean-square between treatments is given by²

$$MS_{\text{Between}} = \frac{SS_{\text{Between}}}{g - 1},$$

the sum-of-squares between treatments is given by

$$SS_{\text{Between}} = \sum_{i=1}^g n_i (\bar{x}_i - \bar{\bar{x}})^2,$$

²The terms MS_{Between} and MS_{Within} are only one set of descriptive labels for the numerator and denominator of the F -ratio test statistic. MS_{Between} is often replaced by either $MS_{\text{Treatment}}$ or MS_{Factor} and MS_{Within} is often replaced by MS_{Error} .

the mean-square within treatments is given by

$$MS_{\text{Within}} = \frac{SS_{\text{Within}}}{N - g},$$

the sum-of-squares within treatments is given by

$$SS_{\text{Within}} = \sum_{i=1}^g \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2,$$

the sum-of-squares total is given by

$$SS_{\text{Total}} = SS_{\text{Between}} + SS_{\text{Within}} = \sum_{i=1}^g \sum_{j=1}^{n_i} (x_{ij} - \bar{\bar{x}})^2,$$

the mean value for the i th of g treatment groups is given by

$$\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij},$$

the grand mean for all g treatment groups combined is given by

$$\bar{\bar{x}} = \frac{1}{N} \sum_{i=1}^g \sum_{j=1}^{n_i} x_{ij},$$

and the total number of observations is

$$N = \sum_{i=1}^g n_i.$$

Under the Neyman–Pearson null hypothesis, $H_0: \mu_1 = \mu_2 = \dots = \mu_g$, test statistic F is asymptotically distributed as Snedecor's F distribution with $\nu_1 = g - 1$ degrees of freedom in the numerator and $\nu_2 = N - g$ degrees of freedom in the denominator. However, if any of the g populations is not normally distributed, then the distribution of test statistic F no longer follows Snedecor's F distribution with $\nu_1 = g - 1$ and $\nu_2 = N - g$ degrees of freedom.

The assumptions underlying Fisher's F -ratio test for multiple independent samples are (1) the observations are independent, (2) the data are random samples from well-defined, normally-distributed populations, and (3) homogeneity of variance; that is, $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_g^2$.

8.2 A Permutation Approach

Now consider a test for multiple independent samples under the Fisher–Pitman permutation model of statistical inference. Under the Fisher–Pitman permutation model there is no null hypothesis specifying population parameters. Instead the null hypothesis simply states that all possible arrangements of the observations occur with equal chance [10]. Also, there is no alternative hypothesis under the permutation model and no specified α level. Moreover, there is no requirement of random sampling, no degrees of freedom, no assumption of normality, and no assumption of homogeneity of variance.

A permutation alternative to the conventional F test for multiple independent samples is easily defined. The permutation test statistic for $g \geq 3$ independent samples is given by

$$\delta = \sum_{i=1}^g C_i \xi_i , \tag{8.1}$$

where $C_i > 0$ is a positive treatment-group weight for $i = 1, \dots, g$,

$$\xi_i = \binom{n_i}{2}^{-1} \sum_{j=1}^{N-1} \sum_{k=j+1}^N \Delta(j, k) \Psi_i(\omega_j) \Psi_i(\omega_k) \tag{8.2}$$

is the average distance-function value for all distinct pairs of objects in sample S_i for $i = 1, \dots, g$,

$$\Delta(j, k) = |x_j - x_k|^v$$

denotes a symmetric distance-function value for a single pair of objects,

$$N = \sum_{i=1}^g n_i ,$$

and $\Psi(\cdot)$ is an indicator function given by

$$\Psi_i(\omega_j) = \begin{cases} 1 & \text{if } \omega_j \in S_i , \\ 0 & \text{otherwise .} \end{cases}$$

Under the Fisher–Pitman permutation model, the null hypothesis simply states that equal probabilities are assigned to each of the

$$M = \frac{N!}{\prod_{i=1}^g n_i!} \tag{8.3}$$

possible, equally-likely allocations of the N objects to the g samples [10]. The probability value associated with an observed value of δ , say δ_o , is the probability under the null hypothesis of observing a value of δ as extreme or more extreme than δ_o . Thus, an exact probability value for δ_o may be expressed as

$$P(\delta \leq \delta_o | H_0) = \frac{\text{number of } \delta \text{ values } \leq \delta_o}{M}. \quad (8.4)$$

When M is large, an approximate probability value for δ may be obtained from a Monte Carlo permutation procedure, where

$$P(\delta \leq \delta_o | H_0) = \frac{\text{number of } \delta \text{ values } \leq \delta_o}{L}$$

and L denotes the number of randomly-sampled test statistic values. Typically, L is set to a large number to ensure accuracy; for example, $L = 1,000,000$ [11].

8.3 The Relationship Between Statistics F and δ

When the null hypothesis under the Neyman–Pearson population model states $H_0: \mu_1 = \mu_2 = \dots = \mu_g$, $v = 2$, and the treatment-group weights are given by

$$C_i = \frac{n_i - 1}{N - g}, \quad i = 1, \dots, g,$$

the functional relationships between test statistic δ and Fisher's F -ratio test statistic are given by

$$\delta = \frac{2SS_{\text{Total}}}{N - g + (g - 1)F} \quad \text{and} \quad F = \frac{2SS_{\text{Total}}}{(g - 1)\delta} - \frac{N - g}{g - 1}, \quad (8.5)$$

where

$$SS_{\text{Total}} = \sum_{i=1}^N x_i^2 - \left(\sum_{i=1}^N x_i \right)^2 / N,$$

and x_i is a univariate measurement score for the i th of N objects. The permutation analogue of the F test is generally known as the Fisher–Pitman permutation test [3].

Because of the relationship between test statistics δ and F , the exact probability values given by

$$P(\delta \leq \delta_o | H_0) = \frac{\text{number of } \delta \text{ values } \leq \delta_o}{M}$$

and

$$P(F \geq F_o | H_0) = \frac{\text{number of } F \text{ values } \geq F_o}{M}$$

are equivalent under the Fisher–Pitman null hypothesis, where δ_o and F_o denote the observed values of δ and F , respectively, and M is the number of possible, equally-likely arrangements of the observed data.

A chance-corrected measure of agreement among the N measurement scores is given by

$$\mathfrak{R} = 1 - \frac{\delta}{\mu_\delta}, \tag{8.6}$$

where μ_δ is the arithmetic average of the M δ test statistic values calculated on all possible arrangements of the observed measurements; that is,

$$\mu_\delta = \frac{1}{M} \sum_{i=1}^M \delta_i. \tag{8.7}$$

Alternatively, in terms of a one-way analysis of variance model, the exact expected value of test statistic δ is a simple function of the total sum-of-squares; that is,

$$\mu_\delta = \frac{2SS_{\text{Total}}}{N - 1}.$$

8.4 Example 1: Test Statistics F and δ

A small example will serve to illustrate the relationship between test statistics F and δ . Consider the example data listed in Table 8.1 with $g = 3$ treatment groups, sample sizes of $n_1 = n_2 = 3, n_3 = 4$, and $N = n_1 + n_2 + n_3 = 3 + 3 + 4 = 10$ total observations. Under the Neyman–Pearson population model with sample sizes $n_1 = n_2 = 3$, and $n_3 = 4$, treatment-group means $\bar{x}_1 = 3, \bar{x}_2 = 4$, and $\bar{x}_3 = 8$, grand mean $\bar{\bar{x}} = 5.30$, estimated population variances $s_1^2 = s_2^2 = 1.00$ and

Table 8.1 Example data for a test of $g = 3$ independent samples with $N = 10$ observations

Treatment group		
1	2	3
2	3	7
3	4	8
4	5	8
		9

$s_3^2 = 0.6667$, the sum-of-squares between treatments is

$$SS_{\text{Between}} = \sum_{i=1}^g n_i (\bar{x}_i - \bar{\bar{x}})^2 = 50.10 ,$$

the sum-of-squares within treatments is

$$SS_{\text{Within}} = \sum_{i=1}^g \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 = 6.00 ,$$

the sum-of-squares total is

$$SS_{\text{Total}} = SS_{\text{Between}} + SS_{\text{Within}} = 50.10 + 6.00 = 56.10 ,$$

the mean-square between treatments is

$$MS_{\text{Between}} = \frac{SS_{\text{Between}}}{g - 1} = \frac{50.10}{3 - 1} = 25.05 ,$$

the mean-square within treatments is

$$MS_{\text{Within}} = \frac{SS_{\text{Within}}}{N - g} = \frac{6.00}{10 - 3} = 0.8571 ,$$

and the observed value of Fisher's F -ratio test statistic is

$$F = \frac{MS_{\text{Between}}}{MS_{\text{Within}}} = \frac{25.05}{0.8571} = 29.2250 .$$

The essential factors, sums of squares (SS), degrees of freedom (df), mean squares (MS), and variance-ratio test statistic (F) are summarized in Table 8.2.

Under the Neyman–Pearson null hypothesis, $H_0: \mu_1 = \mu_2 = \mu_3$, Fisher's F -ratio test statistic is asymptotically distributed as Snedecor's F with $\nu_1 = g - 1$ and $\nu_2 = N - g$ degrees of freedom. With $\nu_1 = g - 1 = 3 - 1 = 2$ and $\nu_2 = N - g = 10 - 3 = 7$ degrees of freedom, the asymptotic probability value of $F = 29.2250$ is $P = 0.4001 \times 10^{-3}$, under the assumptions of normality and homogeneity.

Table 8.2 Source table for the example data listed in Table 8.1

Factor	SS	df	MS	F
Between	50.10	2	25.0500	29.2250
Within	6.00	7	0.8571	
Total	56.10			

8.4.1 An Exact Analysis with $v = 2$

For the first permutation analysis of the example data listed in Table 8.1 let $v = 2$, employing squared Euclidean scaling, and let the treatment-group weights be given by

$$C_i = \frac{n_i - 1}{N - g}, \quad i = 1, \dots, g,$$

for correspondence with Fisher's F -ratio test statistic.

Because there are only

$$M = \frac{N!}{\prod_{i=1}^g n_i!} = \frac{10!}{3! 3! 4!} = 4200$$

possible, equally-likely arrangements in the reference set of all permutations of the $N = 10$ observations listed in Table 8.1, an exact permutation analysis is feasible. While $M = 4200$ arrangements are too many to list, Table 8.3 illustrates the calculation of the ξ , δ , and F values for a small sample of the M possible arrangements of the $N = 10$ observations listed in Table 8.1.

Following Eq. (8.1) on p. 261, the $N = 10$ observations yield $g = 3$ average distance-function values of

$$\xi_i = \xi_2 = 2.00 \quad \text{and} \quad \xi_3 = 1.3333.$$

Alternatively, in terms of a one-way analysis of variance model the average distance-function values are $\xi_1 = 2s_1^2 = 2(1.00) = 2.00$, $\xi_2 = 2s_2^2 = 2(1.00) = 2.00$, and $\xi_3 = 2s_3^2 = 2(0.6667) = 1.3333$.

Following Eq. (8.1) on p. 260, the observed value of the permutation test statistic based on $v = 2$ and treatment-group weights

$$C_i = \frac{n_i - 1}{N - g}, \quad i = 1, 2, 3,$$

is

$$\begin{aligned} \delta = \sum_{i=1}^g C_i \xi_i &= \frac{1}{10 - 3} [(3 - 1)(2.00) + (3 - 1)(2.00) \\ &\quad + (4 - 1)(1.3333)] = 1.7143. \end{aligned}$$

Table 8.3 Sample arrangements of the example data listed in Table 8.1 with associated $\xi_1, \xi_2, \delta,$ and F values

Number	Arrangement	ξ_1	ξ_2	ξ_3	δ	F
1	234 345 7889	2.0000	2.0000	1.3333	1.7143	29.2250
2	234 347 5889	2.0000	8.6667	6.0000	5.6190	6.4839
3	234 347 4889	2.0000	8.6667	9.8333	7.2619	4.4318
4	234 457 3889	2.0000	4.6667	14.6667	8.1905	3.3494
5	234 348 5789	2.0000	14.0000	5.8333	7.0714	4.4333
6	234 358 4789	2.0000	12.6667	9.3333	8.1905	3.3494
7	234 458 3789	2.0000	8.6667	13.8333	8.9762	2.7499
8	234 378 4589	2.0000	14.0000	11.3333	9.4286	2.4500
9	234 478 3589	2.0000	8.6667	15.1667	9.5476	2.3758
10	234 578 3489	2.0000	4.6667	17.3333	9.3333	2.5107
11	234 348 5789	2.0000	14.0000	5.8333	7.0714	4.4333
12	234 358 4789	2.0000	12.6667	9.3333	8.1905	3.3494
13	234 458 3789	2.0000	8.6667	13.8333	8.9762	2.7499
14	234 378 4589	2.0000	14.0000	11.3333	9.4286	2.4500
15	234 478 3589	2.0000	8.6667	15.1667	9.5476	2.3758
16	234 578 3489	2.0000	4.6667	17.3333	9.3333	2.5107
17	234 488 4579	2.0000	10.6667	9.8333	7.8333	3.1894
18	234 488 3579	2.0000	10.6667	13.3333	9.3333	2.5107
19	234 588 3479	2.0000	6.0000	15.1667	8.7857	2.8854
20	234 788 3459	2.0000	0.6667	13.8333	6.6905	4.8851
\vdots	$\vdots \quad \vdots \quad \vdots$	\vdots	\vdots	\vdots	\vdots	\vdots
4199	889 357 2344	0.6667	8.0000	1.8333	3.2619	13.6985
4200	889 457 2343	0.6667	4.6667	1.3333	2.0952	23.2750

Alternatively, in terms of a one-way analysis of variance model the permutation test statistic is

$$\delta = 2MS_{\text{Within}} = 2(0.8571) = 1.7143 .$$

For the example data listed in Table 8.1, the sum of the $N = 10$ observations is

$$\sum_{i=1}^N x_i = 2 + 3 + 4 + 3 + 4 + 5 + 7 + 8 + 8 + 9 = 53 ,$$

the sum of the $N = 10$ squared observations is

$$\sum_{i=1}^N x_i^2 = 2^2 + 3^2 + 4^2 + 3^2 + 4^2 + 5^2 + 7^2 + 8^2 + 8^2 + 9^2 = 337 ,$$

and the total sum-of-squares is

$$\begin{aligned} SS_{\text{Total}} &= \sum_{i=1}^N (x_i - \bar{x})^2 = \sum_{i=1}^N x_i^2 - \left(\sum_{i=1}^N x_i \right)^2 / N \\ &= 337 - (53)^2 / 10 = 56.10, \end{aligned}$$

where \bar{x} denotes the grand mean of all $N = 10$ observations. Then following the expressions given in Eq. (8.5) on p. 262 for test statistics δ and F , the observed value of test statistic δ with respect to test statistic F is

$$\delta = \frac{2SS_{\text{Total}}}{N - g + (g - 1)F} = \frac{2(56.10)}{10 - 3 + (3 - 1)(29.2250)} = 1.7143$$

and the observed value of test statistic F with respect to test statistic δ is

$$F = \frac{2SS_{\text{Total}}}{(g - 1)\delta} - \frac{N - g}{g - 1} = \frac{2(56.10)}{(3 - 1)(1.7143)} - \frac{10 - 3}{3 - 1} = 29.2250.$$

Under the Fisher–Pitman permutation model, the exact probability of an observed δ is the proportion of δ test statistic values computed on all possible, equally-likely arrangements of the $N = 10$ observations listed in Table 8.1 that are equal to or less than the observed value of $\delta = 1.7143$. There are exactly 10 δ test statistic values that are equal to or less than the observed value of $\delta = 1.7143$. If all M arrangements of the $N = 10$ observations listed in Table 8.1 occur with equal chance under the Fisher–Pitman null hypothesis, the exact probability value of $\delta = 1.7143$ computed on all $M = 4200$ arrangements of the observed data with $n_1 = n_2 = 3$ and $n_3 = 4$ preserved for each arrangement is

$$P(\delta \leq \delta_o | H_0) = \frac{\text{number of } \delta \text{ values} \leq \delta_o}{M} = \frac{10}{4200} = 0.2381 \times 10^{-2},$$

where δ_o denotes the observed value of test statistic δ and M is the number of possible, equally-likely arrangements of the $N = 10$ observations listed in Table 8.1.

Alternatively, there are only 10 F values that are larger than the observed value of $F = 29.2250$. Thus, if all arrangements of the observed data occur with equal chance, the exact probability value of $F = 29.2250$ under the Fisher–Pitman null hypothesis is

$$P(F \geq F_o | H_0) = \frac{\text{number of } F \text{ values} \geq F_o}{M} = \frac{10}{4200} = 0.2381 \times 10^{-2},$$

where F_o denotes the observed value of test statistic F .

Following Eq. (8.7) on p. 263, the exact expected value of the $M = 4200$ δ test statistic values under the Fisher–Pitman null hypothesis is

$$\mu_{\delta} = \frac{1}{M} \sum_{i=1}^M \delta_i = \frac{52,360}{4200} = 12.4667 .$$

Alternatively, in terms of a one-way analysis of variance model the exact expected value of test statistic δ is

$$\mu_{\delta} = \frac{2SS_{\text{Total}}}{N - 1} = \frac{2(56.10)}{10 - 1} = 12.4667 .$$

Following Eq. (8.6) on p. 263, the observed chance-corrected measure of effect size is

$$\mathfrak{R} = 1 - \frac{\delta}{\mu_{\delta}} = 1 - \frac{1.7143}{12.4667} = +0.8625 ,$$

indicating approximately 86% within-group agreement above what is expected by chance. Alternatively, in terms of a one-way analysis of variance model the chance-corrected measure of effect size is

$$\begin{aligned} \mathfrak{R} &= 1 - \frac{\delta}{\mu_{\delta}} = 1 - \frac{2MS_{\text{Within}}}{\frac{2SS_{\text{Total}}}{N - 1}} = 1 - \frac{(N - 1)(MS_{\text{Within}})}{SS_{\text{Total}}} \\ &= 1 - \frac{(10 - 1)(0.8571)}{56.10} = +0.8625 . \end{aligned}$$

8.5 Example 2: Measures of Effect Size

Measures of effect size express the practical or clinical significance of differences among multiple independent sample means, as contrasted with the statistical significance of differences. Five measures of effect size are commonly used for determining the magnitude of treatment effects for multiple independent samples: Cohen's \hat{d} , Pearson's η^2 , Kelley's $\hat{\eta}^2$, Hays' $\hat{\omega}_F^2$ for fixed models, and Hays' $\hat{\omega}_R^2$, for random models. Cohen's \hat{d} measure of effect size is given by

$$\hat{d} = \left[\frac{1}{g - 1} \left(\frac{SS_{\text{Between}}}{nMS_{\text{Within}}} \right) \right]^{1/2} = \left[\frac{F}{n} \right]^{1/2} ,$$

where n denotes the common size of each treatment group. Pearson’s η^2 measure of effect size is given by

$$\eta^2 = \frac{SS_{\text{Between}}}{SS_{\text{Total}}} = 1 - \frac{N - g}{F(g - 1) + N - g},$$

which is equivalent to Pearson’s r^2 for a one-way analysis of variance design. Kelley’s “unbiased” correlation ratio is given by³

$$\hat{\eta}^2 = \frac{SS_{\text{Total}} - (N - 1)MS_{\text{Within}}}{SS_{\text{Total}}} = 1 - \frac{N - 1}{F(g - 1) + N - g},$$

which is equivalent to an adjusted or “shrunk” squared multiple correlation coefficient reported by most computer statistical packages and given by

$$\hat{\eta}^2 = R_{\text{adj}}^2 = 1 - \frac{(1 - R^2)(N - 1)}{N - p - 1},$$

where R^2 is the squared product-moment multiple correlation coefficient and p is the number of predictors. Hays’ $\hat{\omega}_F^2$ measure of effect size for a fixed-effects analysis of variance model is given by

$$\hat{\omega}_F^2 = \frac{SS_{\text{Between}} - (g - 1)MS_{\text{Within}}}{SS_{\text{Total}} + MS_{\text{Within}}} = 1 - \frac{N}{(F - 1)(g - 1) + N}.$$

Hays’ $\hat{\omega}_R^2$ measure of effect size for a random-effects analysis of variance model is given by

$$\hat{\omega}_R^2 = \frac{MS_{\text{Between}} - MS_{\text{Within}}}{MS_{\text{Between}} + (n - 1)MS_{\text{Within}}} = 1 - \frac{n}{F + n - 1},$$

where n denotes the common size of each treatment group. Mielke and Berry’s \mathfrak{R} chance-corrected measure of effect size is given by

$$\mathfrak{R} = 1 - \frac{\delta}{\mu_\delta},$$

where δ is defined in Eq. (8.1) on p. 261 and μ_δ is the exact expected value of δ under the Fisher–Pitman null hypothesis given by

$$\mu_\delta = \frac{1}{M} \sum_{i=1}^M \delta_i,$$

³It is well known that Kelley’s correlation ratio is not unbiased, but since the title of Truman Kelley’s 1935 article was “An unbiased correlation ratio measure,” the label has persisted.

where, for a test of $g \geq 3$ independent samples, the number of possible, equally-likely arrangements of the observed data is given by

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

For the example data listed in Table 8.1 on p. 263 for $N = 10$ observations, Cohen's \hat{d} measure of effect size is⁴

$$\hat{d} = \left[\frac{1}{g-1} \left(\frac{SS_{\text{Between}}}{\bar{n}MS_{\text{Within}}} \right) \right]^{1/2} = \left[\frac{F}{\bar{n}} \right]^{1/2} = \left[\frac{29.2250}{3.3333} \right]^{1/2} = \pm 2.9610 .$$

Pearson's r^2 measure of effect size is usually labeled as η^2 when reported with an analysis of variance. For the example data listed in Table 8.1, η^2 is

$$\begin{aligned} \eta^2 &= \frac{SS_{\text{Between}}}{SS_{\text{Total}}} = 1 - \frac{N-g}{F(g-1) + N-g} \\ &= 1 - \frac{10-3}{(29.2250)(3-1) + 10-3} = 0.8930 , \end{aligned}$$

Kelley's $\hat{\eta}^2$ measure of effect size is

$$\begin{aligned} \hat{\eta}^2 &= \frac{SS_{\text{Total}} - (N-1)MS_{\text{Within}}}{SS_{\text{Total}}} = 1 - \frac{N-1}{F(g-1) + N-g} \\ &= 1 - \frac{10-1}{(29.2250)(3-1) + 10-3} = 0.8625 , \end{aligned}$$

Hays' $\hat{\omega}_F^2$ measure of effect size for a fixed-effects analysis of variance model is

$$\begin{aligned} \hat{\omega}_F^2 &= \frac{SS_{\text{Between}} - (g-1)MS_{\text{Within}}}{SS_{\text{Total}} + MS_{\text{Within}}} = 1 - \frac{N}{(F-1)(g-1) + N} \\ &= 1 - \frac{10}{(29.2250-1)(3-1) + 10} = 0.8495 , \end{aligned}$$

⁴Since the sizes of the treatment groups are not equal, the average value of $\bar{n} = 3.3333$ is used for both Cohen's \hat{d} measure of effect size and Hays' $\hat{\omega}_R^2$ measure of effect size for a random-effects model. In cases where the treatment-group sizes differ greatly, a weighted average recommended by Haggard is often adopted [6].

Hays' $\hat{\omega}_R^2$ measure of effect size for a random-effects analysis of variance model is⁵

$$\begin{aligned}\hat{\omega}_R^2 &= \frac{MS_{\text{Between}} - MS_{\text{Within}}}{MS_{\text{Between}} + (\bar{n} - 1)MS_{\text{Within}}} = 1 - \frac{\bar{n}}{F + \bar{n} - 1} \\ &= 1 - \frac{3.3333}{29.2250 + 3.3333 - 1} = 0.8944 ,\end{aligned}$$

and Mielke and Berry's \mathfrak{R} chance-corrected measure of effect size is

$$\mathfrak{R} = 1 - \frac{\delta}{\mu_\delta} = 1 - \frac{1.7143}{12.4667} = +0.8625 ,$$

where the exact expected value of test statistic δ under the Fisher–Pitman null hypothesis is

$$\mu_\delta = \frac{1}{M} \sum_{i=1}^M \delta_i = \frac{52,360}{4200} = 12.4667 .$$

It can easily be shown that Mielke and Berry's \mathfrak{R} chance-corrected measure of effect size is identical to Kelley's $\hat{\eta}^2$ measure of effect size for a one-way, completely-randomized analysis of variance design, under the Neyman–Pearson population model.

8.5.1 Comparisons of Effect Size Measures

In this section the various measures of effect size are compared and contrasted. Because Pearson's r^2 and η^2 are equivalent and Kelley's $\hat{\eta}^2$ and Mielke and Berry's \mathfrak{R} are equivalent for multi-sample designs, only η^2 and \mathfrak{R} are utilized for the comparisons. The functional relationships between Cohen's \hat{d} measure of effect size and Pearson's η^2 (r^2) measure of effect size for $g \geq 3$ independent samples are given by

$$\hat{d} = \left[\frac{\eta^2(N - g)}{n(g - 1)(1 - \eta^2)} \right]^{1/2} \quad \text{and} \quad \eta^2 = 1 - \frac{N - g}{n\hat{d}^2(g - 1) + N - g} , \quad (8.8)$$

where n denotes the common treatment-group size. The relationships between Cohen's \hat{d} measure of effect size and Mielke and Berry's \mathfrak{R} ($\hat{\eta}^2$) chance-corrected

⁵For a one-way completely-randomized analysis of variance, a fixed-effects model and a random-effects model yield the same F -ratio, but measures of effect size can differ under the two models.

measure of effect size are given by

$$\hat{d} = \left[\frac{\mathfrak{R}(N - g) + g - 1}{n(g - 1)(1 - \mathfrak{R})} \right]^{1/2} \quad \text{and} \quad \mathfrak{R} = 1 - \frac{N - 1}{n\hat{d}^2(g - 1) + N - g}. \quad (8.9)$$

The relationships between Cohen's \hat{d} measure of effect size and Hays' $\hat{\omega}_F^2$ measure of effect size for a fixed-effects model are given by

$$\hat{d} = \left[\frac{(N - g + 1)\hat{\omega}_F^2 + g - 1}{n(g - 1)(1 - \hat{\omega}_F^2)} \right]^{1/2} \quad (8.10)$$

and

$$\hat{\omega}_F^2 = 1 - \frac{N}{(n\hat{d}^2 - 1)(g - 1) + N}. \quad (8.11)$$

The relationships between Cohen's \hat{d} measure of effect size and Hays' $\hat{\omega}_R^2$ measure of effect size for a random-effects model are given by

$$\hat{d} = \left[\frac{\hat{\omega}_R^2(n - 1) + 1}{n(1 - \hat{\omega}_R^2)} \right]^{1/2} \quad \text{and} \quad \hat{\omega}_R^2 = 1 - \frac{n}{n(\hat{d}^2 + 1) - 1}. \quad (8.12)$$

The relationships between Pearson's η^2 (r^2) measure of effect size and Mielke and Berry's \mathfrak{R} ($\hat{\eta}^2$) measure of effect size are given by

$$\eta^2 = 1 - \frac{(N - g)(1 - \mathfrak{R})}{N - 1} \quad \text{and} \quad \mathfrak{R} = 1 - \frac{(N - 1)(1 - \eta^2)}{N - g}. \quad (8.13)$$

The relationships between Pearson's η^2 (r^2) measure of effect size and Hays' $\hat{\omega}_F^2$ measure of effect size for a fixed-effects model are given by

$$\eta^2 = \frac{(N - g + 1)\hat{\omega}_F^2 + g - 1}{N + \hat{\omega}_F^2 - 1} \quad (8.14)$$

and

$$\hat{\omega}_F^2 = \frac{\eta^2(N - 1) - g + 1}{N - \eta^2 - g + 1}. \quad (8.15)$$

The relationships between Pearson's η^2 (r^2) measure of effect size and Hays' $\hat{\omega}_R^2$ measure of effect size for a random-effects model are given by

$$\eta^2 = 1 - \frac{(N - g)(1 - \hat{\omega}_R^2)}{(g - 1)[\hat{\omega}_R^2(n - 1) + 1] + (N - g)(1 - \hat{\omega}_R^2)} \quad (8.16)$$

and

$$\hat{\omega}_R^2 = \frac{\eta^2(N - 1) - g + 1}{(N - g)\eta^2 + (g - 1)(1 - \eta^2)(n - 1)}. \quad (8.17)$$

The relationships between Mielke and Berry's \mathfrak{R} ($\hat{\eta}^2$) measure of effect size and Hays' $\hat{\omega}_F^2$ measure of effect size for a fixed-effects model are given by

$$\mathfrak{R} = \frac{N\hat{\omega}_F^2}{N + \hat{\omega}_F^2 - 1} \quad \text{and} \quad \hat{\omega}_F^2 = \frac{\mathfrak{R}(N - 1)}{N - \mathfrak{R}}. \quad (8.18)$$

The relationships between Mielke and Berry's \mathfrak{R} ($\hat{\eta}^2$) measure of effect size and Hays' $\hat{\omega}_R^2$ measure of effect size for a random-effects model are given by

$$\mathfrak{R} = 1 - \frac{(N - 1)(1 - \hat{\omega}_R^2)}{n\hat{\omega}_R^2(g - 1) + (N - 1)(1 - \hat{\omega}_R^2)} \quad (8.19)$$

and

$$\hat{\omega}_R^2 = \frac{\hat{\eta}^2(N - 1)}{N\mathfrak{R} - 1 + (1 - \mathfrak{R})[n(g - 1) + 1]}. \quad (8.20)$$

And the relationships between Hays' $\hat{\omega}_F^2$ measure of effect size for a fixed-effects model and Hays' $\hat{\omega}_R^2$ measure of effect size for a random-effects model are given by

$$\hat{\omega}_F^2 = \frac{n\hat{\omega}_R^2(g - 1)}{n\hat{\omega}_R^2 + N(1 - \hat{\omega}_R^2)} \quad \text{and} \quad \hat{\omega}_R^2 = \frac{N\hat{\omega}_F^2}{N\hat{\omega}_F^2 - n(g - 1)(1 - \hat{\omega}_F^2)}. \quad (8.21)$$

8.5.2 Example Comparisons of Effect Size Measures

In this section comparisons of Cohen's \hat{d} , Pearson's η^2 , Mielke and Berry's \mathfrak{R} , Hays' $\hat{\omega}_F^2$, and Hays' $\hat{\omega}_R^2$ measures of effect size are illustrated with the example data listed in Table 8.1 on p. 263 with $n_1 = n_2 = 3$, $n_3 = 4$, and $N = n_1 + n_2 + n_3 = 3 + 3 + 4 = 10$ observations. Because the treatment-group sizes are unequal, the

ns in the equations for Cohen's \hat{d} and Hays' $\hat{\omega}_R^2$ are replaced with a simple average; that is, $\bar{n} = (3 + 3 + 4)/3 = 3.3333$.

Given the example data listed in Table 8.1 and following the expressions given in Eq. (8.8) for Cohen's \hat{d} measure of effect size and Pearson's η^2 (r^2) measure of effect size, the observed value for Cohen's \hat{d} measure of effect size with respect to the observed value of Pearson's η^2 (r^2) measure of effect size is

$$\hat{d} = \left[\frac{\eta^2(N - g)}{\bar{n}(g - 1)(1 - \eta^2)} \right]^{1/2} = \left[\frac{(0.8930)(10 - 3)}{(3.3333)(3 - 1)(1 - 0.8930)} \right]^{1/2} = \pm 2.9610$$

and the observed value for Pearson's η^2 (r^2) measure of effect size with respect to the observed value of Cohen's \hat{d} measure of effect size is

$$\begin{aligned} \eta^2 &= 1 - \frac{N - g}{\bar{n}\hat{d}^2(g - 1) + N - g} \\ &= 1 - \frac{10 - 3}{(3.3333)(2.9610)^2(3 - 1) + 10 - 3} = 0.8930 . \end{aligned}$$

Following the expressions given in Eq. (8.9) for Cohen's \hat{d} measure of effect size and Mielke and Berry's \mathfrak{R} ($\hat{\eta}^2$) measure of effect size, the observed value for Cohen's \hat{d} measure of effect size with respect to the observed value of Mielke and Berry's \mathfrak{R} ($\hat{\eta}^2$) measure of effect size is

$$\begin{aligned} \hat{d} &= \left[\frac{\mathfrak{R}(N - g) + g - 1}{\bar{n}(g - 1)(1 - \mathfrak{R})} \right]^{1/2} \\ &= \left[\frac{0.8625(10 - 3) + 3 - 1}{(3.3333)(3 - 1)(1 - 0.8625)} \right]^{1/2} = \pm 2.9610 \end{aligned}$$

and the observed value for Mielke and Berry's \mathfrak{R} ($\hat{\eta}^2$) measure of effect size with respect to the observed value of Cohen's \hat{d} measure of effect size is

$$\begin{aligned} \mathfrak{R} &= 1 - \frac{N - 1}{\bar{n}\hat{d}^2(g - 1) + N - g} \\ &= 1 - \frac{10 - 1}{(3.3333)(2.9610)^2(3 - 1) + 10 - 3} = +0.8625 . \end{aligned}$$

Following the expressions given in Eqs. (8.10) and (8.11) for Cohen's \hat{d} measure of effect size and Hays' $\hat{\omega}_F^2$ measure of effect size for a fixed-effects model, the observed value for Cohen's \hat{d} measure of effect size with respect to the observed

value of Hays' $\hat{\omega}_F^2$ measure of effect size is

$$\begin{aligned}\hat{d} &= \left[\frac{(N - g + 1)\hat{\omega}_F^2 + g - 1}{\bar{n}(g - 1)(1 - \hat{\omega}_F^2)} \right]^{1/2} \\ &= \left[\frac{(10 - 3 + 1)(0.8495) + 3 - 1}{(3.3333)(3 - 1)(1 - 0.8495)} \right]^{1/2} = \pm 2.9610\end{aligned}$$

and the observed value for Hays' $\hat{\omega}_F^2$ measure of effect size with respect to the observed value of Cohen's \hat{d} measure of effect size is

$$\begin{aligned}\hat{\omega}_F^2 &= 1 - \frac{N}{(\bar{n}\hat{d}^2 - 1)(g - 1) + N} \\ &= 1 - \frac{10}{[(3.3333)(2.9610)^2 - 1](3 - 1) + 10} = 0.8495.\end{aligned}$$

Following the expressions given in Eq. (8.12) for Cohen's \hat{d} measure of effect size and Hays' $\hat{\omega}_R^2$ measure of effect size for a random-effects model, the observed value for Cohen's \hat{d} measure of effect size with respect to the observed value of Hays' $\hat{\omega}_R^2$ measure of effect size is

$$\hat{d} = \left[\frac{\hat{\omega}_R^2(\bar{n} - 1) + 1}{\bar{n}(1 - \hat{\omega}_R^2)} \right]^{1/2} = \left[\frac{(0.8944)(3.3333 - 1) + 1}{(3.3333)(1 - 0.8944)} \right]^{1/2} = \pm 2.9610$$

and the observed value of Hays' $\hat{\omega}_R^2$ measure of effect size with respect to the observed value of Cohen's \hat{d} measure of effect size is

$$\hat{\omega}_R^2 = 1 - \frac{\bar{n}}{\bar{n}(\hat{d}^2 + 1) - 1} = 1 - \frac{3.3333}{(3.3333)[(2.9610)^2 + 1] - 1} = 0.8944.$$

Following the expressions given in Eq. (8.13) for Pearson's η^2 (r^2) measure of effect size and Mielke and Berry's \mathfrak{R} ($\hat{\eta}^2$) measure of effect size, the observed value for Pearson's η^2 (r^2) measure of effect size with respect to the observed value of Mielke and Berry's \mathfrak{R} ($\hat{\eta}^2$) measure of effect size is

$$\eta^2 = 1 - \frac{(N - g)(1 - \mathfrak{R})}{N - 1} = 1 - \frac{(10 - 3)(1 - 0.8625)}{10 - 1} = 0.8930$$

and the observed value for Mielke and Berry's \mathfrak{R} ($\hat{\eta}^2$) measure of effect size with respect to the observed value of Pearson's η^2 (r^2) measure of effect size is

$$\mathfrak{R} = 1 - \frac{(N - 1)(1 - \eta^2)}{N - g} = 1 - \frac{(10 - 1)(1 - 0.8930)}{10 - 3} = +0.8625.$$

Following the expressions given in Eqs. (8.14) and (8.15) for Pearson's η^2 (r^2) measure of effect size and Hays' $\hat{\omega}_F^2$ measure of effect size for a fixed-effects model, the observed value for Pearson's η^2 (r^2) measure of effect size with respect to the observed value of Hays' $\hat{\omega}_F^2$ measure of effect size is

$$\eta^2 = \frac{(N - g + 1)\hat{\omega}_F^2 + g - 1}{N + \hat{\omega}_F^2 - 1} = \frac{(10 - 3 + 1)(0.8495) + 3 - 1}{10 + 0.8495 - 1} = 0.8930$$

and the observed value for Hays' $\hat{\omega}_F^2$ measure of effect size with respect to the observed value of Pearson's η^2 (r^2) measure of effect size is

$$\hat{\omega}_F^2 = \frac{\eta^2(N - 1) - g + 1}{N - \eta^2 - g + 1} = \frac{(0.8930)(10 - 1) - 3 + 1}{10 - 0.8930 - 3 + 1} = 0.8495 .$$

Following the expressions given in Eqs. (8.16) and (8.17) for Pearson's η^2 (r^2) measure of effect size and Hays' $\hat{\omega}_R^2$ measure of effect size for a random-effects model, the observed value for Pearson's η^2 (r^2) measure of effect size with respect to the observed value of Hays' $\hat{\omega}_R^2$ measure of effect size is

$$\begin{aligned} \eta^2 &= 1 - \frac{(N - g)(1 - \hat{\omega}_R^2)}{(g - 1)[\hat{\omega}_R^2(\bar{n} - 1) + 1] + (N - g)(1 - \hat{\omega}_R^2)} \\ &= 1 - \frac{(10 - 3)(1 - 0.8944)}{(3 - 1)[(0.8944)(3.3333 - 1) + 1] + (10 - 3)(1 - 0.8944)} \\ &= 0.8930 \end{aligned}$$

and the observed value for Hays' $\hat{\omega}_R^2$ measure of effect size with respect to the observed value of Pearson's η^2 (r^2) measure of effect size is

$$\begin{aligned} \hat{\omega}_R^2 &= \frac{\eta^2(N - 1) - g + 1}{(N - g)\eta^2 + (g - 1)(1 - \eta^2)(\bar{n} - 1)} \\ &= \frac{0.8930(10 - 1) - 3 + 1}{(10 - 3)(0.8930) + (3 - 1)(1 - 0.8930)(3.3333 - 1)} = 0.8944 . \end{aligned}$$

Following the expressions given in Eq. (8.18) for Mielke and Berry's \mathfrak{R} ($\hat{\eta}^2$) measure of effect size and Hays' $\hat{\omega}_F^2$ measure of effect size for a fixed-effects model, the observed value for Mielke and Berry's \mathfrak{R} ($\hat{\eta}^2$) measure of effect size with respect to the observed value of Hays' $\hat{\omega}_F^2$ measure of effect size is

$$\mathfrak{R} = \frac{N\hat{\omega}_F^2}{N + \hat{\omega}_F^2 - 1} = \frac{(10)(0.8495)}{10 + 0.8495 - 1} = +0.8625$$

and the observed value for Hays' $\hat{\omega}_F^2$ measure of effect size with respect to the observed value of Mielke and Berry's \mathfrak{R} ($\hat{\eta}^2$) measure of effect size is

$$\hat{\omega}_F^2 = \frac{\mathfrak{R}(N-1)}{N-\mathfrak{R}} = \frac{(0.8625)(10-1)}{10-0.8625} = 0.8495 .$$

Following the expressions given in Eqs. (8.19) and (8.20) for Mielke and Berry's \mathfrak{R} ($\hat{\eta}^2$) measure of effect size and Hays' $\hat{\omega}_R^2$ measure of effect size for a random-effects model, the observed value for Mielke and Berry's \mathfrak{R} ($\hat{\eta}^2$) measure of effect size with respect to the observed value of Hays' $\hat{\omega}_R^2$ measure of effect size is

$$\begin{aligned} \mathfrak{R} &= 1 - \frac{(N-1)(1-\hat{\omega}_R^2)}{\bar{n}\hat{\omega}_R^2(g-1) + (N-1)(1-\hat{\omega}_R^2)} \\ &= 1 - \frac{(10-1)(1-0.8944)}{(3.3333)(0.8944)(3-1) + (10-1)(1-0.8944)} = +0.8625 \end{aligned}$$

and the observed value for Hays' $\hat{\omega}_R^2$ measure of effect size with respect to the observed value for Mielke and Berry's \mathfrak{R} ($\hat{\eta}^2$) measure of effect size is

$$\begin{aligned} \hat{\omega}_R^2 &= \frac{\hat{\eta}^2(N-1)}{N\mathfrak{R}-1 + (1-\mathfrak{R})[\bar{n}(g-1) + 1]} \\ &= \frac{(0.8625)(10-1)}{(10)(0.8625) - 1 + (1-0.8625)[(3.3333)(3-1) + 1]} = 0.8944 . \end{aligned}$$

Following the expressions given in Eq. (8.21) for Hays' $\hat{\omega}_F^2$ measure of effect size for a fixed-effects model and Hays' $\hat{\omega}_R^2$ measure of effect size for a random-effects model, the observed value for Hays' $\hat{\omega}_F^2$ measure of effect size with respect to the observed value of Hays' $\hat{\omega}_R^2$ measure of effect size is

$$\begin{aligned} \hat{\omega}_F^2 &= \frac{\bar{n}\hat{\omega}_R^2(g-1)}{\bar{n}\hat{\omega}_R^2(g-1) + N(1-\hat{\omega}_R^2)} \\ &= \frac{(3.3333)(0.8944)(3-1)}{(3.3333)(0.8944)(3-1) + (10)(1-0.8944)} = 0.8495 \end{aligned}$$

and the observed value for Hays' $\hat{\omega}_R^2$ measure of effect size with respect to the observed value of Hays' $\hat{\omega}_F^2$ measure of effect size is

$$\begin{aligned} \hat{\omega}_R^2 &= \frac{N\hat{\omega}_F^2}{N\hat{\omega}_F^2 + \bar{n}(g-1)(1-\hat{\omega}_F^2)} \\ &= \frac{(10)(0.8495)}{(10)(0.8495) + (3.3333)(3-1)(1-0.8495)} = 0.8944 . \end{aligned}$$

8.6 Example 3: Analyses with $v = 2$ and $v = 1$

For a third example of tests of differences among $g \geq 3$ independent samples, consider the example data set given in Table 8.4 with $g = 4$ treatment groups, sample sizes of $n_1 = n_2 = n_3 = n_4 = 7$, and $N = 28$ total observations. Under the Neyman–Pearson population model with sample sizes $n_1 = n_2 = n_3 = n_4 = 7$, treatment-group means $\bar{x}_1 = 20.4286$, $\bar{x}_2 = 20.8571$, $\bar{x}_3 = 9.1429$, and $\bar{x}_4 = 14.1429$, grand mean $\bar{\bar{x}} = 16.1429$, estimated population variances $s_1^2 = 27.9524$, $s_2^2 = 35.4762$, and $s_3^2 = s_4^2 = 8.8095$, the sum-of-squares between treatments is

$$SS_{\text{Between}} = \sum_{i=1}^g n_i (\bar{x}_i - \bar{\bar{x}})^2 = 655.1429 ,$$

the sum-of-squares within treatments is

$$SS_{\text{Within}} = \sum_{i=1}^g \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 = 486.2857 ,$$

the sum-of-squares total is

$$SS_{\text{Total}} = SS_{\text{Between}} + SS_{\text{Within}} = 655.1429 + 486.2857 = 1141.4286 ,$$

the mean-square between treatments is

$$MS_{\text{Between}} = \frac{SS_{\text{Between}}}{g - 1} = \frac{655.1429}{4 - 1} = 218.3810 ,$$

the mean-square within treatments is

$$MS_{\text{Within}} = \frac{SS_{\text{Within}}}{N - g} = \frac{486.28571}{28 - 4} = 20.2619 ,$$

Table 8.4 Example data for a test of $g = 4$ independent samples with $N = 28$ observations

Treatment group			
1	2	3	4
15	24	10	15
23	14	5	13
18	15	8	10
16	19	13	17
25	30	6	18
29	26	10	11
17	18	12	15

Table 8.5 Source table for the data listed in Table 8.4

Factor	SS	df	MS	F
Between	655.1429	3	218.3810	10.7779
Within	486.1429	24	20.2619	
Total	1141.4286			

and the observed value of Fisher’s F -ratio test statistic is

$$F = \frac{MS_{\text{Between}}}{MS_{\text{Within}}} = \frac{218.3810}{20.2619} = 10.7779 .$$

The essential factors, sums of squares (SS), degrees of freedom (df), mean squares (MS), and variance-ratio test statistic (F) are summarized in Table 8.5.

Under the Neyman–Pearson null hypothesis, $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$, Fisher’s F -ratio test statistic is asymptotically distributed as Snedecor’s F with $\nu_1 = g - 1$ and $\nu_2 = N - g$ degrees of freedom. With $\nu_1 = g - 1 = 4 - 1 = 3$ and $\nu_2 = N - g = 28 - 4 = 24$ degrees of freedom, the asymptotic probability value of $F = 10.7778$ is $P = 0.1122 \times 10^{-3}$, under the assumptions of normality and homogeneity.

8.6.1 A Monte Carlo Analysis with $v = 2$

For the first analysis of the example data listed in Table 8.4 on p. 278 under the Fisher–Pitman permutation model let $v = 2$, employing squared Euclidean scaling, and let the treatment-group weights be given by

$$C_i = \frac{n_i - 1}{N - g} , \quad i = 1, \dots, g ,$$

for correspondence with Fisher’s F -ratio test statistic.

Because there are

$$M = \frac{N!}{g \prod_{i=1}^g n_i!} = \frac{28!}{7! 7! 7! 7!} = 472,518,347,558,400$$

possible, equally-likely arrangements in the reference set of all permutations of the $N = 28$ observations listed in Table 8.4, an exact permutation analysis is not possible and a Monte Carlo analysis is required.

Following Eq. (8.2) on p. 261, the $N = 28$ observations yield $g = 4$ average distance-function values of

$$\xi_i = 55.9048 , \quad \xi_2 = 70.9524 , \quad \text{and} \quad \xi_3 = \xi_4 = 17.6190 .$$

Alternatively, in terms of a one-way analysis of variance model the average distance-function values are $\xi_1 = 2s_1^2 = 2(27.9524) = 55.9048$, $\xi_2 = 2s_2^2 = 2(34.4762) = 70.9524$, $\xi_3 = 2s_3^2 = 2(8.8095) = 17.6190$, and $\xi_4 = 2s_4^2 = 2(8.8095) = 17.6190$.

Following Eq. (8.1) on p. 261, the observed value of the permutation test statistic based on $v = 2$ and treatment-group weights

$$C_i = \frac{n_i - 1}{N - g}, \quad i = 1, \dots, 4,$$

is

$$\begin{aligned} \delta = \sum_{i=1}^g C_i \xi_i &= \frac{7-1}{28-4} (55.9048 + 70.9524 \\ &\quad + 17.6190 + 17.6190) = 40.5238. \end{aligned}$$

Alternatively, in terms of a one-way analysis of variance model the permutation test statistic is

$$\delta = 2MS_{\text{Within}} = 2(20.2619) = 40.5238.$$

For the example data listed in Table 8.4, the sum of the $N = 28$ observations is

$$\sum_{i=1}^N x_i = 15 + 23 + 18 + \dots + 11 + 15 = 452,$$

the sum of the $N = 28$ squared observations is

$$\sum_{i=1}^N x_i^2 = 15^2 + 23^2 + 18^2 + \dots + 11^2 + 15^2 = 8438,$$

and the total sum-of-squares is

$$\begin{aligned} SS_{\text{Total}} &= \sum_{i=1}^N (x_i - \bar{\bar{x}})^2 = \sum_{i=1}^N x_i^2 - \left(\sum_{i=1}^N x_i \right)^2 / N \\ &= 8438 - (452)^2 / 28 = 1141.4286, \end{aligned}$$

where $\bar{\bar{x}}$ denotes the grand mean of all $N = 28$ observations.

Then following the expressions given in Eq. (8.5) on p. 262 for test statistics δ and F , the observed value for test statistic δ with respect to the observed value of test statistic F is

$$\delta = \frac{2SS_{\text{Total}}}{N - g + (g - 1)F} = \frac{2(1141.4286)}{28 - 4 + (4 - 1)(10.7779)} = 40.5238$$

and the observed value of test statistic F with respect to the observed value of test statistic δ is

$$F = \frac{2SS_{\text{Total}}}{(g - 1)\delta} - \frac{N - g}{g - 1} = \frac{2(1141.4286)}{(4 - 1)(40.5238)} - \frac{28 - 4}{4 - 1} = 10.7779 .$$

Under the Fisher–Pitman permutation model, the Monte Carlo probability of an observed δ is the proportion of δ test statistic values computed on the randomly-selected, equally-likely arrangements of the $N = 28$ observations listed in Table 8.4 that are equal to or less than the observed value of $\delta = 40.5238$. There are exactly 138 δ test statistic values that are equal to or less than the observed value of $\delta = 40.5238$. If all M arrangements of the $N = 28$ observations listed in Table 8.4 occur with equal chance under the Fisher–Pitman null hypothesis, the Monte Carlo probability value of $\delta = 40.5238$ computed on $L = 1,000,000$ random arrangements of the observed data with $n_1 = n_2 = n_3 = n_4 = 7$ preserved for each arrangement is

$$P(\delta \leq \delta_o) = \frac{\text{number of } \delta \text{ values } \leq \delta_o}{L} = \frac{138}{1,000,000} = 0.1380 \times 10^{-3} ,$$

where δ_o denotes the observed value of test statistic δ and L is the number of randomly-selected, equally-likely arrangements of the $N = 28$ observations listed in Table 8.4.

In terms of a one-way analysis of variance model, there are only 138 F values that are larger than the observed value of $F = 10.7779$. Thus, if all arrangements of the observed data occur with equal chance, the exact probability value of $F = 10.7779$ under the Fisher–Pitman null hypothesis is

$$P(F \geq F_o) = \frac{\text{number of } F \text{ values } \geq F_o}{L} = \frac{138}{1,000,000} = 0.1380 \times 10^{-3} ,$$

where F_o denotes the observed value of test statistic F and L is the number of random, equally-likely arrangements of the example data listed in Table 8.4.

Following Eq. (8.7) on p. 263, the exact expected value of the $M = 4200$ δ test statistic values under the Fisher–Pitman null hypothesis is

$$\mu_\delta = \frac{1}{M} \sum_{i=1}^M \delta_i = \frac{39,951,568,041,566,987}{472,518,347,558,400} = 84.5503 .$$

Alternatively, in terms of a one-way analysis of variance model the exact expected value of test statistic δ under the Fisher–Pitman null hypothesis is

$$\mu_\delta = \frac{2SS_{\text{Total}}}{N-1} = \frac{2(1141.4286)}{28-1} = 84.5503 .$$

Following Eq. (8.6) on p. 263, the observed chance-corrected measure of effect size is

$$\mathfrak{R} = 1 - \frac{\delta}{\mu_\delta} = 1 - \frac{40.5238}{84.5503} = +0.5207 ,$$

indicating approximately 52% within-group agreement above what is expected by chance. Alternatively, in terms of a one-way analysis of variance model, the observed chance-corrected measure of effect size is

$$\mathfrak{R} = 1 - \frac{(N-1)(MS_{\text{Within}})}{SS_{\text{Total}}} = 1 - \frac{(28-1)(20.2619)}{1141.4286} = +0.5207 .$$

Alternatively, in terms of Fisher’s F -ratio test statistic the chance-corrected measure of effect size is

$$\mathfrak{R} = 1 - \frac{N-1}{F(g-1) + N-g} = 1 - \frac{28-1}{10.7779(4-1) + 28-4} = +0.5207 .$$

8.6.2 Measures of Effect Size

For the example data listed in Table 8.4, Cohen’s \hat{d} measure of effect size is

$$\hat{d} = \left[\frac{1}{g-1} \left(\frac{SS_{\text{Between}}}{nMS_{\text{Within}}} \right) \right]^{1/2} = \left[\frac{1}{4-1} \left(\frac{655.1429}{(7)(20.2619)} \right) \right]^{1/2} = \pm 1.2408 ,$$

Pearson’s η^2 (r^2) measure of effect size is

$$\eta^2 = \frac{SS_{\text{Between}}}{SS_{\text{Total}}} = \frac{655.1429}{1141.4286} = 0.5740 ,$$

Kelley’s $\hat{\eta}^2$ measure of effect size is

$$\begin{aligned} \hat{\eta}^2 &= \frac{SS_{\text{Total}} - (N-1)MS_{\text{Within}}}{SS_{\text{Total}}} \\ &= \frac{1141.4286 - (28-1)(20.2619)}{1141.4286} = 0.5207 , \end{aligned}$$

Hays' $\hat{\omega}_F^2$ measure of effect size for a fixed-effects model is

$$\begin{aligned}\hat{\omega}_F^2 &= \frac{SS_{\text{Between}} - (g - 1)MS_{\text{Within}}}{SS_{\text{Total}} + MS_{\text{Within}}} \\ &= \frac{655.1429 - (4 - 1)(20.2619)}{1141.4286 + 20.2619} = 0.5116 ,\end{aligned}$$

Hays' $\hat{\omega}_R^2$ measure of effect size for a random-effects model is

$$\begin{aligned}\hat{\omega}_R^2 &= \frac{MS_{\text{Between}} - MS_{\text{Within}}}{MS_{\text{Between}} + (n - 1)MS_{\text{Within}}} \\ &= \frac{655.1429 - 20.2619}{655.1429 + (7 - 1)(20.2619)} = 0.8174 ,\end{aligned}$$

and the observed chance-corrected measure of effect size is

$$\mathfrak{R} = 1 - \frac{\delta}{\mu_\delta} = 1 - \frac{40.5238}{84.5503} = +0.5207 ,$$

indicating approximately 52% within-group agreement above what is expected by chance.

8.6.3 A Monte Carlo Analysis with $v = 1$

Consider a second analysis of the example data listed in Table 8.4 on p. 278 under the Fisher–Pitman permutation model with $v = 1$ and treatment-group weights

$$C_i = \frac{n_i - 1}{N - g} , \quad i = 1, \dots, g .$$

For $v = 1$, the average distance-function values for the $g = 4$ treatment groups are

$$\xi_1 = 6.2857 , \quad \xi_2 = 7.2381 , \quad \text{and} \quad \xi_3 = \xi_4 = 3.6190 ,$$

respectively, and the observed permutation test statistic is

$$\begin{aligned}\delta &= \sum_{i=1}^g C_i \xi_i \\ &= \left(\frac{7 - 1}{28 - 4} \right) (6.2857 + 7.2381 + 3.6190 + 3.6190) = 5.1905 .\end{aligned}$$

Because there are

$$M = \frac{N!}{\prod_{i=1}^g n_i!} = \frac{28!}{7! 7! 7! 7!} = 472,518,347,558,400$$

possible, equally-likely arrangements in the reference set of all permutations of the $N = 28$ observations listed in Table 8.4, an exact permutation analysis is impossible and a Monte Carlo permutation analysis is required. Under the Fisher–Pitman permutation model, the Monte Carlo probability of an observed δ is the proportion of δ test statistic values computed on the randomly-selected, equally-likely arrangements of the $N = 28$ observations listed in Table 8.4 that are equal to or less than the observed value of $\delta = 5.1905$. There are exactly 204 δ test statistic values that are equal to or less than the observed value of $\delta = 5.1905$. If all M arrangements of the $N = 28$ observations listed in Table 8.4 occur with equal chance under the Fisher–Pitman null hypothesis, the Monte Carlo probability value of $\delta = 5.1905$ computed on $L = 1,000,000$ random arrangements of the observed data with $n_1 = n_2 = n_3 = n_4 = 7$ preserved for each arrangement is

$$P(\delta \leq \delta_o | H_0) = \frac{\text{number of } \delta \text{ values} \leq \delta_o}{L} = \frac{204}{1,000,000} = 0.2040 \times 10^{-3},$$

where δ_o denotes the observed value of test statistic δ and L is the number of randomly-selected, equally-likely arrangements of the $N = 28$ observations listed in Table 8.4. No comparison is made with Fisher’s F -ratio test statistic as F is undefined for ordinary Euclidean scaling.

For the example data listed in Table 8.4, the exact expected value of test statistic δ under the Fisher–Pitman null hypothesis is

$$\mu_\delta = \frac{1}{M} \sum_{i=1}^M \delta_i = \frac{3,497,628,060,462,033}{472,518,347,558,400} = 7.4021 \quad (8.22)$$

and the observed chance-corrected measure of effect size is

$$\Re = 1 - \frac{\delta}{\mu_\delta} = 1 - \frac{5.1905}{7.4021} = +0.2988,$$

indicating approximately 30% within-group agreement above what is expected by chance. No comparisons are made with Cohen’s \hat{d} , Pearson’s η^2 (r^2), Kelley’s $\hat{\eta}^2$, Hays’ $\hat{\omega}_F^2$, or Hays’ $\hat{\omega}_R^2$ conventional measures of effect size as \hat{d} , η^2 , $\hat{\eta}^2$, $\hat{\omega}_F^2$, and $\hat{\omega}_R^2$ are undefined for ordinary Euclidean scaling.

8.6.4 The Effects of Extreme Values

To illustrate the robustness to the inclusion of extreme values of ordinary Euclidean scaling with $v = 1$, consider the example data listed in Table 8.4 on p. 278 with one alteration. The seventh (last) observation in Group 4 in Table 8.4 has been increased from $x_{7,4} = 15$ to $x_{7,4} = 75$, as shown in Table 8.6. Under the Neyman–Pearson population model with sample sizes $n_1 = n_2 = n_3 = n_4 = 7$, treatment-group means $\bar{x}_1 = 20.4286$, $\bar{x}_2 = 20.8571$, $\bar{x}_3 = 9.1429$, and $\bar{x}_4 = 22.7143$, grand mean $\bar{\bar{x}} = 18.2857$, estimated population variances $s_1^2 = 27.9524$, $s_2^2 = 35.4762$, $s_3^2 = 8.8095$, and $s_4^2 = 540.2381$, the sum-of-squares between treatments is

$$SS_{\text{Between}} = \sum_{i=1}^g n_i (\bar{x}_i - \bar{\bar{x}})^2 = 800.8571 ,$$

the sum-of-squares within treatments is

$$SS_{\text{Within}} = \sum_{i=1}^g \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 = 3674.8571 ,$$

the sum-of-squares total is

$$SS_{\text{Total}} = SS_{\text{Between}} + SS_{\text{Within}} = 800.8571 + 3674.8571 = 4475.7142 ,$$

the mean-square between treatments is

$$MS_{\text{Between}} = \frac{SS_{\text{Between}}}{g - 1} = \frac{655.1429}{4 - 1} = 266.9524 ,$$

the mean-square within treatments is

$$MS_{\text{Within}} = \frac{SS_{\text{Within}}}{N - g} = \frac{486.28571}{28 - 4} = 153.1190 ,$$

Table 8.6 Example data for a test of $g = 4$ independent samples with $N = 28$ observations and one extreme value, $x_{7,4} = 75$

Treatment group			
1	2	3	4
15	24	10	15
23	14	5	13
18	15	8	10
16	19	13	17
25	30	6	18
29	26	10	11
17	18	12	75

Table 8.7 Source table for the data listed in Table 8.6

Factor	SS	df	MS	F
Between	800.8571	3	266.9524	1.7434
Within	3674.8571	24	153.1190	
Total	4475.7142			

and the observed value of Fisher's F -ratio test statistic is

$$F = \frac{MS_{\text{Between}}}{MS_{\text{Within}}} = \frac{266.9524}{153.1190} = 1.7434 .$$

The essential factors, sums of squares (SS), degrees of freedom (df), mean squares (MS), and variance-ratio test statistic (F) are summarized in Table 8.7.

Under the Neyman–Pearson null hypothesis, $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$, Fisher's F -ratio test statistic is asymptotically distributed as Snedecor's F with $\nu_1 = g - 1$ and $\nu_2 = N - g$ degrees of freedom. With $\nu_1 = g - 1 = 4 - 1 = 3$ and $\nu_2 = N - g = 28 - 4 = 24$ degrees of freedom, the asymptotic probability value of $F = 1.7434$ is $P = 0.1849$, under the assumptions of normality and homogeneity. The original F -ratio test statistic value with observation $x_{7,4} = 15$ was $F = 10.7779$ with an asymptotic probability value of $P = 0.1122 \times 10^{-3}$, yielding a difference between the two probability values of

$$\Delta P = 0.1849 - 0.1122 \times 10^{-3} = 0.1848 .$$

8.6.5 A Monte Carlo Analysis with $v = 2$

For the first analysis of the example data listed in Table 8.6 on p. 285 under the Fisher–Pitman permutation model let $v = 2$, employing squared Euclidean scaling, and let the treatment-group weights be given by

$$C_i = \frac{n_i - 1}{N - g} , \quad i = 1, \dots, g ,$$

for correspondence with Fisher's F -ratio test statistic.

Because there are

$$M = \frac{N!}{g \prod_{i=1}^g n_i!} = \frac{28!}{7! 7! 7! 7!} = 472,518,347,558,400$$

possible, equally-likely arrangements in the reference set of all permutations of the $N = 28$ observations listed in Table 8.6, an exact permutation analysis is not possible and a Monte Carlo analysis is required.

Following Eq. (8.2) on p. 261, the $N = 28$ observations yield $g = 4$ average distance-function values of

$$\xi_1 = 55.9048, \quad \xi_2 = 70.9524, \quad \xi_3 = 17.6190, \quad \text{and} \quad \xi_4 = 1080.4762.$$

Alternatively, under an analysis of variance model, $\xi_1 = 2s_1^2 = 2(27.9524) = 55.9048$, $\xi_2 = 2s_2^2 = 2(35.4762) = 70.9524$, $\xi_3 = 2s_3^2 = 2(8.8095) = 17.6190$, and $\xi_4 = 2s_4^2 = 2(540.2381) = 1080.4762$.

Following Eq. (8.1) on p. 261, the observed value of the permutation test statistic based on $v = 2$ and treatment-group weights

$$C_i = \frac{n_i - 1}{N - g}, \quad i = 1, \dots, 4,$$

is

$$\begin{aligned} \delta = \sum_{i=1}^g C_i \xi_i &= \frac{7 - 1}{28 - 4} (55.9048 + 70.9524 \\ &\quad + 17.6190 + 1080.4762) = 306.2381. \end{aligned}$$

Under the Fisher–Pitman permutation model, the Monte Carlo probability of an observed δ is the proportion of δ test statistic values computed on the randomly-selected, equally-likely arrangements of the $N = 28$ observations listed in Table 8.6 that are equal to or less than the observed value of $\delta = 306.2381$. There are exactly 128,239 δ test statistic values that are equal to or less than the observed value of $\delta = 306.2381$. If all M arrangements of the $N = 28$ observations listed in Table 8.6 occur with equal chance under the Fisher–Pitman null hypothesis, the Monte Carlo probability value of $\delta = 306.2381$ computed on $L = 1,000,000$ random arrangements of the observed data with $n_1 = n_2 = n_3 = n_4 = 7$ preserved for each arrangement is

$$P(\delta \leq \delta_o | H_0) = \frac{\text{number of } \delta \text{ values} \leq \delta_o}{L} = \frac{128,239}{1,000,000} = 0.1282,$$

where δ_o denotes the observed value of test statistic δ and L is the number of randomly-selected, equally-likely arrangements of the $N = 28$ observations listed in Table 8.6. For comparison, the original value of test statistic δ based on $v = 2$ with observation $x_{7,4} = 15$ was $\delta = 40.5238$ with a Monte Carlo probability value of $P = 0.1380 \times 10^{-3}$, yielding a difference between the two probability values of

$$\Delta_P = 0.1282 - 0.1380 \times 10^{-3} = 0.1281.$$

8.6.6 A Monte Carlo Analysis with $v = 1$

For the second analysis of the example data listed in Table 8.6 on p. 285 under the Fisher–Pitman permutation model let $v = 1$, employing ordinary Euclidean scaling, and let the treatment-group weights be given by

$$C_i = \frac{n_i - 1}{N - g}, \quad i = 1, \dots, g.$$

Setting $v = 1$ can be expected to reduce the outsized effect of extreme value $x_{7,4} = 75$.

Because there are

$$M = \frac{N!}{\frac{g}{\prod_{i=1}^g n_i!}} = \frac{28!}{7! 7! 7! 7!} = 472,518,347,558,400$$

possible, equally-likely arrangements in the reference set of all permutations of the $N = 28$ observations listed in Table 8.6, an exact permutation analysis is not possible and a Monte Carlo analysis is required.

Following Eq. (8.2) on p. 261, the $N = 28$ observations yield $g = 4$ average distance-function values of

$$\xi_1 = 6.2857, \quad \xi_2 = 7.2381, \quad \xi_3 = 3.6190, \quad \text{and} \quad \xi_4 = 20.2857.$$

Following Eq. (8.1) on p. 261, the observed value of the permutation test statistic based on $v = 1$ and treatment-group weights

$$C_i = \frac{n_i - 1}{N - g}, \quad i = 1, \dots, 4,$$

is

$$\delta = \sum_{i=1}^g C_i \xi_i = \frac{7 - 1}{28 - 4} (6.2857 + 7.2381 + 3.6190 + 20.2857) = 9.3571.$$

Under the Fisher–Pitman permutation model, the exact probability of an observed δ is the proportion of δ test statistic values computed on the randomly-selected, equally-likely arrangements of the $N = 28$ observations listed in Table 8.6 that are equal to or less than the observed value of $\delta = 9.3571$. There are exactly 1960 δ test statistic values that are equal to or less than the observed value of $\delta = 9.3571$. If all M arrangements of the $N = 28$ observations listed in Table 8.6 occur with equal chance, the Monte Carlo probability value of $\delta = 9.3571$

computed on $L = 1,000,000$ random arrangements of the observed data with $n_1 = n_2 = n_3 = n_4 = 7$ preserved for each arrangement is

$$P(\delta \leq \delta_o | H_0) = \frac{\text{number of } \delta \text{ values } \leq \delta_o}{L} = \frac{1960}{1,000,000} = 0.1960 \times 10^{-2},$$

where δ_o denotes the observed value of δ and L is the number of randomly-selected, equally-likely arrangements of the $N = 28$ observations listed in Table 8.6.

The original value of test statistic δ based on $v = 1$ with observation $x_{7,4} = 15$ was $\delta = 5.1905$ with a Monte Carlo probability value of $P = 0.2040 \times 10^{-3}$, yielding a difference between the two probability values of only

$$\Delta_P = 0.1960 \times 10^{-2} - 0.2040 \times 10^{-3} = 0.1756 \times 10^{-2}.$$

Multi-sample permutation tests based on ordinary Euclidean scaling with $v = 1$ tend to be relatively robust with respect to extreme values when compared with permutation tests based on squared Euclidean scaling with $v = 2$.

8.7 Example 4: Exact and Monte Carlo Analyses

For a fourth, larger example of tests for differences among $g \geq 3$ independent samples, consider the example data given in Table 8.8 with $g = 4$ treatment groups, sample sizes of $n_1 = n_2 = 3, n_3 = 4, n_4 = 5$, and $N = n_1 + n_2 + n_3 + n_4 = 3 + 3 + 4 + 5 = 15$ total observations. Under the Neyman–Pearson population model with sample sizes $n_1 = n_2 = 3, n_3 = 4$, and $n_4 = 5$, treatment-group means $\bar{x}_1 = 11.00, \bar{x}_2 = 12.00, \bar{x}_3 = 13.50$, and $\bar{x}_4 = 19.00$, grand mean $\bar{\bar{x}} = 14.5333$, estimated population variances $s_1^2 = s_2^2 = 1.00, s_3^2 = 1.6667$, and $s_4^2 = 62.50$, the sum-of-squares between treatments is

$$SS_{\text{Between}} = \sum_{i=1}^g n_i (\bar{x}_i - \bar{\bar{x}})^2 = 160.7333,$$

Table 8.8 Example data for a test of $g = 4$ independent samples with $N = 15$ observations

Treatment group			
1	2	3	4
10	11	12	14
11	12	13	15
12	13	14	16
		15	17
			33

Table 8.9 Source table for the data listed in Table 8.8

Factor	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>
Between	160.7333	3	53.5778	2.2755
Within	259.0000	11	23.5455	
Total	419.7333			

the sum-of-squares within treatments is

$$SS_{\text{Within}} = \sum_{i=1}^g \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 = 259.00 ,$$

the sum-of-squares total is

$$SS_{\text{Total}} = SS_{\text{Between}} + SS_{\text{Within}} = 160.7333 + 259.00 = 419.7333 ,$$

the mean-square between treatments is

$$MS_{\text{Between}} = \frac{SS_{\text{Between}}}{g - 1} = \frac{160.7333}{4 - 1} = 53.5778 ,$$

the mean-square within treatments is

$$MS_{\text{Within}} = \frac{SS_{\text{Within}}}{N - g} = \frac{259.00}{15 - 4} = 23.5455 ,$$

and the observed value of Fisher's *F*-ratio test statistic is

$$F = \frac{MS_{\text{Between}}}{MS_{\text{Within}}} = \frac{53.5778}{23.5455} = 2.2755 .$$

The essential factors, sums of squares (*SS*), degrees of freedom (*df*), mean squares (*MS*), and variance-ratio test statistic (*F*) are summarized in Table 8.9.

Under the Neyman–Pearson null hypothesis, $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$, Fisher's *F*-ratio test statistic is asymptotically distributed as Snedecor's *F* with $\nu_1 = g - 1$ and $\nu_2 = N - g$ degrees of freedom. With $\nu_1 = g - 1 = 4 - 1 = 3$ and $\nu_2 = N - g = 15 - 4 = 11$ degrees of freedom, the asymptotic probability value of $F = 2.2755$ is $P = 0.1366$, under the assumptions of normality and homogeneity.

8.7.1 A Permutation Analysis with $v = 2$

For the first analysis of the example data listed in Table 8.8 under the Fisher–Pitman permutation model let $v = 2$, employing squared Euclidean scaling, and let the

treatment-group weighting functions be given by

$$C_i = \frac{n_i - 1}{N - g}, \quad i = 1, \dots, g,$$

for correspondence with Fisher’s F -ratio test statistic.

Because there are

$$M = \frac{N!}{g \prod_{i=1}^g n_i!} = \frac{15!}{3! 3! 4! 5!} = 12,612,600$$

possible, equally-likely arrangements in the reference set of all permutations of the $N = 15$ observations listed in Table 8.8, an exact permutation analysis is not practical and a Monte Carlo analysis is utilized.

Following Eq. (8.2) on p. 261, the $N = 15$ observations yield $g = 4$ average distance-function values of

$$\xi_1 = \xi_2 = 2.00, \quad \xi_3 = 3.3333, \quad \text{and} \quad \xi_4 = 125.00.$$

Alternatively, in terms of a one-way analysis of variance model the average distance-function values are $\xi_1 = 2s_1^2 = 2(1.00) = 2.00$, $\xi_2 = 2s_2^2 = 2(1.00) = 2.00$, $\xi_3 = 2s_3^2 = 2(1.667) = 3.3333$, and $\xi_4 = 2s_4^2 = 2(62.50) = 125.00$.

Following Eq. (8.1) on p. 261, the observed value of the permutation test statistic based on $v = 2$ and treatment-group weights

$$C_i = \frac{n_i - 1}{N - g}, \quad i = 1, \dots, 4,$$

is

$$\begin{aligned} \delta = \sum_{i=1}^g C_i \xi_i &= \frac{1}{15 - 4} [(3 - 1)(2.00) + (3 - 1)(2.00) \\ &\quad + (4 - 1)(3.3333) + (5 - 1)(125.00)] = 47.0909. \end{aligned}$$

Alternatively, in terms of a one-way analysis of variance model the permutation test statistic is

$$\delta = 2MS_{\text{Within}} = 2(23.5455) = 47.0909.$$

For the example data listed in Table 8.8, the sum of the $N = 15$ observations is

$$\sum_{i=1}^N x_i = 10 + 11 + 12 + \dots + 17 + 33 = 218,$$

the sum of the $N = 15$ squared observations is

$$\sum_{i=1}^N x_i^2 = 10^2 + 11^2 + 12^2 + \cdots + 17^2 + 33^2 = 3588 ,$$

and the total sum-of-squares is

$$\begin{aligned} SS_{\text{Total}} &= \sum_{i=1}^N (x_i - \bar{x})^2 = \sum_{i=1}^N d_i^2 - \left(\sum_{i=1}^N d_i \right)^2 / N \\ &= 3588 - (218)^2 / 15 = 419.7333 , \end{aligned}$$

where \bar{x} denotes the grand mean of all $N = 15$ observations. Then following the expressions given in Eq. (8.5) on p. 262 for test statistics δ and F , the observed value for test statistic δ with respect to the observed value of test statistic F is

$$\delta = \frac{2SS_{\text{Total}}}{N - g + (g - 1)F} = \frac{2(419.7333)}{15 - 4 + (4 - 1)(2.2755)} = 47.0909$$

and the observed value for test statistic F with respect to the observed value of test statistic δ is

$$F = \frac{2SS_{\text{Total}}}{(g - 1)\delta} - \frac{N - g}{g - 1} = \frac{2(419.7333)}{(4 - 1)(47.0909)} - \frac{15 - 4}{4 - 1} = 2.2755 .$$

Under the Fisher–Pitman permutation model, the Monte Carlo probability of an observed δ is the proportion of δ test statistic values computed on the randomly-selected, equally-likely arrangements of the $N = 15$ observations listed in Table 8.8 that are equal to or less than the observed value of $\delta = 47.0909$. There are exactly 53,200 δ test statistic values that are equal to or less than the observed value of $\delta = 47.0909$. If all M arrangements of the $N = 15$ observations listed in Table 8.8 occur with equal chance under the Fisher–Pitman null hypothesis, the Monte Carlo probability value of $\delta = 47.0909$ computed on $L = 1,000,000$ randomly-selected arrangements of the observed data with $n_1 = n_2 = 3 = n_3 = 4$, and $n_4 = 5$ preserved for each arrangement is

$$P(\delta \leq \delta_o | H_0) = \frac{\text{number of } \delta \text{ values } \leq \delta_o}{L} = \frac{53,200}{1,000,000} = 0.0532 ,$$

where δ_o denotes the observed value of test statistic δ and L is the number of randomly-selected, equally-likely arrangements of the $N = 15$ observations listed in Table 8.8.

Alternatively, in terms of a one-way analysis of variance model, there are 53,200 F values that are equal to or greater than the observed value of $F = 2.2755$. Thus, if

all arrangements of the observed data occur with equal chance, the exact probability value of $F = 2.2755$ under the Fisher–Pitman null hypothesis is

$$P(F \geq F_o | H_0) = \frac{\text{number of } F \text{ values} \geq F_o}{L} = \frac{53,200}{1,000,000} = 0.0532 ,$$

where F_o denotes the observed value of test statistic F .

Following Eq. (8.7) on p. 263, the exact expected value of the $M = 12,612,600$ δ test statistic values under the Fisher–Pitman null hypothesis is

$$\mu_\delta = \frac{1}{M} \sum_{i=1}^M \delta_i = \frac{756,275,456}{12,612,600} = 59.9619 .$$

In terms of a one-way analysis of variance model the exact expected value of test statistic δ is

$$\mu_\delta = \frac{2SS_{\text{Total}}}{N - 1} = \frac{2(419.7333)}{15 - 1} = 59.9619 .$$

Following Eq. (8.6) on p. 263, the observed chance-corrected measure of effect size is

$$\mathfrak{R} = 1 - \frac{\delta}{\mu_\delta} = 1 - \frac{47.0909}{59.9619} = +0.2147 ,$$

indicating approximately 21% within-group agreement above what is expected by chance. Alternatively, in terms of a one-way analysis of variance model, the observed measure of effect size is

$$\mathfrak{R} = 1 - \frac{(N - 1)(MS_{\text{Within}})}{SS_{\text{Total}}} = 1 - \frac{(15 - 1)(23.5455)}{419.7333} = +0.2147 .$$

8.7.2 Measures of Effect Size

For the example data listed in Table 8.8 on p. 289, the average treatment-group size is

$$\bar{n} = \frac{1}{g} \sum_{i=1}^g n_i = \frac{3 + 3 + 4 + 5}{4} = 3.75 ,$$

Cohen's \hat{d} measure of effect size is

$$\begin{aligned}\hat{d} &= \left[\frac{1}{g-1} \left(\frac{SS_{\text{Between}}}{\bar{n}MS_{\text{Within}}} \right) \right]^{1/2} \\ &= \left[\frac{1}{4-1} \left(\frac{160.7333}{(3.75)(23.5455)} \right) \right]^{1/2} = \pm 0.7336 ,\end{aligned}$$

Pearson's η^2 (r^2) measure of effect size is

$$\eta^2 = \frac{SS_{\text{Between}}}{SS_{\text{Total}}} = \frac{160.7333}{419.7333} = 0.3829 ,$$

Kelley's $\hat{\eta}^2$ measure of effect size is

$$\begin{aligned}\hat{\eta}^2 &= \frac{SS_{\text{Total}} - (N-1)MS_{\text{Within}}}{SS_{\text{Total}}} \\ &= \frac{419.7333 - (15-1)(23.5455)}{419.7333} = 0.2147 ,\end{aligned}$$

Hays' $\hat{\omega}_F^2$ measure of effect size for a fixed-effects model is

$$\begin{aligned}\hat{\omega}_F^2 &= \frac{SS_{\text{Between}} - (g-1)MS_{\text{Within}}}{SS_{\text{Total}} + MS_{\text{Within}}} \\ &= \frac{160.7333 - (4-1)(23.5455)}{419.7333 + 23.5455} = 0.2033 ,\end{aligned}$$

Hays' $\hat{\omega}_R^2$ measure of effect size for a random-effects model is

$$\begin{aligned}\hat{\omega}_R^2 &= \frac{MS_{\text{Between}} - MS_{\text{Within}}}{MS_{\text{Between}} + (\bar{n}-1)MS_{\text{Within}}} \\ &= \frac{53.5777 - 23.5455}{53.5777 + (3.75)(23.5455)} = 0.2117 ,\end{aligned}$$

and Mielke and Berry's \mathfrak{R} chance-corrected measure of effect size is

$$\mathfrak{R} = 1 - \frac{\delta}{\mu_\delta} = 1 - \frac{47.0909}{56.9619} = +0.2147 ,$$

indicating approximately 21% within-group agreement above what is expected by chance.

8.7.3 An Exact Analysis with $v = 2$

While an exact permutation analysis with $M = 12,612,600$ possible arrangements of the observed data may be impractical, it is not impossible. An exact analysis of the $N = 15$ observations listed in Table 8.8 on p. 289 under the Fisher–Pitman permutation model yields $g = 4$ average distance-function values of

$$\xi_1 = \xi_2 = 2.00, \quad \xi_3 = 3.3333, \quad \text{and} \quad \xi_4 = 125.00.$$

The observed value of the permutation test statistic based on $v = 2$ and treatment-group weights

$$C_i = \frac{n_i - 1}{N - g}, \quad i = 1, \dots, 4,$$

is

$$\begin{aligned} \delta = \sum_{i=1}^g C_i \xi_i &= \frac{1}{15 - 4} [(3 - 1)(2.00) + (3 - 1)(2.00) \\ &\quad + (4 - 1)(3.3333) + (5 - 1)(125.00)] = 47.0909. \end{aligned}$$

Under the Fisher–Pitman permutation model, the exact probability of an observed δ is the proportion of δ test statistic values computed on all possible, equally-likely arrangements of the $N = 15$ observations listed in Table 8.8 that are equal to or less than the observed value of $\delta = 47.0909$. There are exactly 673,490 δ test statistic values that are equal to or less than the observed value of $\delta = 47.0909$. If all M arrangements of the $N = 15$ observations listed in Table 8.8 occur with equal chance under the Fisher–Pitman null hypothesis, the exact probability value of $\delta = 47.0909$ computed on the $M = 12,612,600$ possible arrangements of the observed data with $n_1 = n_2 = 3 = n_3 = 4$, and $n_4 = 5$ preserved for each arrangement is

$$P(\delta \leq \delta_o | H_0) = \frac{\text{number of } \delta \text{ values } \leq \delta_o}{M} = \frac{673,490}{12,612,600} = 0.0534,$$

where δ_o denotes the observed value of test statistic δ and M is the number of possible, equally-likely arrangements of the $N = 15$ observations listed in Table 8.8.

Carrying the Monte Carlo probability value based on $L = 1,000,000$ random arrangements and the exact probability value based on $M = 12,612,600$ possible arrangements to a few extra decimal places allows for a more direct comparison of the Monte Carlo and exact permutation approaches. The Monte Carlo approximate probability value and the corresponding exact probability value to six decimal places are

$$P = 0.053242 \quad \text{and} \quad P = 0.053398,$$

respectively. The difference between the two probability values is only

$$\Delta_P = 0.053398 - 0.053242 = 0.000156 ,$$

demonstrating the efficiency and accuracy of a Monte Carlo approach for permutation methods when L is large and the exact probability value is not too small. In general, $L = 1,000,000$ random arrangements of the observed data is sufficient to ensure three decimal places of accuracy [11].

8.7.4 A Monte Carlo Analysis with $v = 1$

Consider a second analysis of the example data listed in Table 8.8 on p. 289 under the Fisher–Pitman permutation model with $v = 1$ and treatment-group weights

$$C_i = \frac{n_i - 1}{N - g} , \quad i = 1, \dots, g .$$

For $v = 1$, employing ordinary Euclidean scaling between the observations, thereby reducing the effects of any extreme values, the average distance-function values for the $g = 4$ treatment groups are

$$\xi_1 = \xi_2 = 1.3333 , \quad \xi_3 = 1.6667 , \quad \text{and} \quad \xi_4 = 8.00 ,$$

respectively, and the observed permutation test statistic is

$$\begin{aligned} \delta = \sum_{i=1}^g C_i \xi_i &= \left(\frac{1}{15 - 4} \right) (3 - 1)(1.3333) + (3 - 1)(1.3333) \\ &\quad + (4 - 1)(1.6667) + (5 - 1)(8.00) = 3.8485 . \end{aligned}$$

Because there are

$$M = \frac{N!}{\prod_{i=1}^g n_i!} = \frac{15!}{3! 3! 4! 5!} = 12,612,600$$

possible, equally-likely arrangements in the reference set of all permutations of the $N = 28$ observations listed in Table 8.8, a Monte Carlo permutation analysis is recommended.

Under the Fisher–Pitman permutation model, the Monte Carlo probability of an observed δ is the proportion of δ test statistic values computed on the randomly-selected, equally-likely arrangements of the $N = 15$ observations listed in Table 8.8

that are equal to or less than the observed value of $\delta = 3.8485$. There are exactly 18,000 δ test statistic values that are equal to or less than the observed value of $\delta = 3.8485$. If all M arrangements of the $N = 15$ observations listed in Table 8.8 occur with equal chance under the Fisher–Pitman null hypothesis, the Monte Carlo probability value of $\delta = 3.8485$ computed on $L = 1,000,000$ random arrangements of the observed data with $n_1 = n_2 = 3$, $n_3 = 4$, and $n_4 = 5$ preserved for each arrangement is

$$P(\delta \leq \delta_o | H_0) = \frac{\text{number of } \delta \text{ values } \leq \delta_o}{L} = \frac{18,000}{1,000,000} = 0.0180 ,$$

where δ_o denotes the observed value of test statistic δ and L is the number of randomly-selected, equally-likely arrangements of the $N = 15$ observations listed in Table 8.8.

For comparison, the approximate Monte Carlo probability value based on $v = 2$, $L = 1,000,000$, and

$$C_i = \frac{n_i - 1}{N - g} , \quad i = 1, \dots, g ,$$

is $P = 0.0532$. The difference between the two probability values, $P = 0.0180$ and $P = 0.0532$, is due to the single extreme value of $x_{5,4} = 33$ in the fourth treatment group. No comparison is made with Fisher's F -ratio test statistic as F is undefined for ordinary Euclidean scaling.

For the example data listed in Table 8.8 on p. 289, the exact expected value of the $M = 12,612,600$ δ test statistic values under the Fisher–Pitman null hypothesis is

$$\mu_\delta = \frac{1}{M} \sum_{i=1}^M \delta_i = \frac{59,579,400}{12,612,600} = 4.7238 \quad (8.23)$$

and the observed chance-corrected measure of effect size is

$$\mathfrak{N} = 1 - \frac{\delta}{\mu_\delta} = 1 - \frac{3.8485}{4.7238} = +0.1853 ,$$

indicating approximately 19% within-group agreement above what is expected by chance. No comparisons are made with Cohen's \hat{d} , Pearson's η^2 (r^2), Kelley's $\hat{\eta}^2$, Hays' $\hat{\omega}_F^2$, or Hays' $\hat{\omega}_R^2$ conventional measures of effect size as \hat{d} , η^2 , $\hat{\eta}^2$, $\hat{\omega}_F^2$, and $\hat{\omega}_R^2$ are undefined for ordinary Euclidean scaling.

8.7.5 An Exact Analysis with $v = 1$

An exact permutation analysis of the observations listed in Table 8.8 with $v = 1$ yields $g = 4$ average distance-function values of

$$\xi_1 = \xi_2 = 1.3333, \quad \xi_3 = 1.6667, \quad \text{and} \quad \xi_4 = 8.00.$$

The observed value of the permutation test statistic based on $v = 1$ and treatment-group weights

$$C_i = \frac{n_i - 1}{N - g}, \quad i = 1, \dots, 4,$$

is

$$\begin{aligned} \delta = \sum_{i=1}^g C_i \xi_i &= \frac{1}{15 - 4} [(3 - 1)(1.3333) + (3 - 1)(1.3333) \\ &\quad + (4 - 1)(1.6667) + (5 - 1)(8.00)] = 3.8485. \end{aligned}$$

Under the Fisher–Pitman permutation model, the exact probability of an observed δ is the proportion of δ test statistic values computed on all possible, equally-likely arrangements of the $N = 15$ observations listed in Table 8.8 that are equal to or less than the observed value of $\delta = 3.8485$. There are exactly 225,720 δ test statistic values that are equal to or less than the observed value of $\delta = 3.8485$. If all M arrangements of the $N = 15$ observations listed in Table 8.8 occur with equal chance under the Fisher–Pitman null hypothesis, the exact probability value of $\delta = 3.8485$ computed on the $M = 12,612,600$ possible arrangements of the observed data with $n_1 = n_2 = 3$, $n_3 = 4$, and $n_4 = 5$ preserved for each arrangement is

$$P(\delta \leq \delta_o | H_0) = \frac{\text{number of } \delta \text{ values } \leq \delta_o}{M} = \frac{225,720}{12,612,600} = 0.0179,$$

where δ_o denotes the observed value of test statistic δ and M is the number of possible, equally-likely arrangements of the $N = 15$ observations listed in Table 8.8.

The exact expected value of the $M = 12,612,600$ δ test statistic values under the Fisher–Pitman null hypothesis is

$$\mu_\delta = \frac{1}{M} \sum_{i=1}^M \delta_i = \frac{59,579,400}{12,612,600} = 4.7238$$

and the observed chance-corrected measure of effect size is

$$\mathfrak{R} = 1 - \frac{\delta}{\mu_\delta} = 1 - \frac{3.8485}{4.7238} = +0.1853,$$

indicating approximately 19% within-group agreement above what is expected by chance. No comparisons are made with Cohen's \hat{d} , Pearson's η^2 (r^2), Kelley's $\hat{\eta}^2$, Hays' $\hat{\omega}_F^2$, or Hays' $\hat{\omega}_R^2$ conventional measures of effect size as \hat{d} , η^2 , $\hat{\eta}^2$, $\hat{\omega}_F^2$, and $\hat{\omega}_R^2$ are undefined for ordinary Euclidean scaling.

Finally, note the effect of a single extreme value ($x_{4,5} = 33$) in Treatment 4 in the analysis based on ordinary Euclidean scaling with $\nu = 1$, compared with the analysis based on squared Euclidean scaling with $\nu = 2$. In the analysis based on $\nu = 2$, the value for the fourth average distance-function value was $\xi_4 = 125.00$, but in the analysis based on $\nu = 1$, ξ_4 was reduced to only $\xi_4 = 8.00$. Also, in the analysis based on $\nu = 2$ the exact probability value was $P = 0.0534$, but in the analysis based on $\nu = 1$ the exact probability value was only $P = 0.0179$, a reduction of approximately 66%. For comparison, the asymptotic probability value of $F = 2.2755$ with $\nu_1 = g - 1 = 4 - 1 = 3$ and $\nu_2 = N - g = 15 - 4 = 11$ degrees of freedom was $P = 0.1366$.

8.8 Example 5: Rank-Score Permutation Analyses

In many research applications it becomes necessary to analyze rank-score data, typically because the required parametric assumptions of normality and homogeneity cannot be met. Consequently, the raw scores are often converted to rank scores and analyzed under a less-restrictive model. While it is never necessary to convert raw scores to rank scores under the Fisher–Pitman permutation model, sometimes the observed data are simply collected as rank scores. Thus, this fifth example serves merely to demonstrate the relationship between a g -sample test of rank-score observations under the population model and the same test under the permutation model. The conventional approach to univariate rank-score data for multiple independent samples under the Neyman–Pearson population model is the Kruskal–Wallis g -sample rank-sum test. As Kruskal and Wallis explained, the rank-sum test stemmed from two statistical methods: rank transformations of the original raw scores and permutations of the rank-order statistics [12].

8.8.1 The Kruskal–Wallis Rank-Sum Test

Consider g random samples of possibly different sizes and denote the size of the i th sample by n_i for $i = 1, \dots, g$. Let

$$N = \sum_{i=1}^g n_i$$

denote the total number of observations, assign rank 1 to the smallest of the N observations, rank 2 to the next smallest observation, continuing to the largest observation that is assigned rank N , and let R_i denote the sum of the rank scores in the i th sample, $i = 1, \dots, g$. If there are no tied rank scores, the Kruskal–Wallis g -sample rank-sum test statistic is given by

$$H = \frac{12}{N(N + 1)} \sum_{i=1}^g \frac{R_i}{n_i} - 3(N + 1) . \tag{8.24}$$

When $g = 2$, H is equivalent to the Wilcoxon [25], Festinger [5], Mann–Whitney [15], Haldane–Smith [7], and van der Reyden [24] two-sample rank-sum tests.

For an example analysis of g -sample rank-score data, consider the rank scores listed in Table 8.10 with $g = 3$ samples, $n_1 = n_2 = n_3 = 6$, $N = 18$, and no tied rank scores.

The conventional Kruskal–Wallis g -sample rank-sum test on the $N = 18$ rank scores listed in Table 8.10 yields an observed test statistic of

$$\begin{aligned} H &= \frac{12}{N(N + 1)} \sum_{i=1}^g \frac{R_i}{n_i} - 3(N + 1) \\ &= \frac{12}{18(18 + 1)} \left[\frac{(63)^2}{6} + \frac{(30)^2}{6} + \frac{(78)^2}{6} \right] - 3(18 + 1) = 7.0526 , \end{aligned}$$

where test statistic H is asymptotically distributed as Pearson’s chi-squared under the Neyman–Pearson null hypothesis with $g - 1$ degrees of freedom as $N \rightarrow \infty$. Under the Neyman–Pearson null hypothesis with $g - 1 = 3 - 1 = 2$ degrees of freedom, the observed value of $H = 7.0526$ yields an asymptotic probability value of $P = 0.0294$, under the assumption of normality.

Table 8.10 Ranking of $g = 3$ with $n_1 = n_2 = n_3 = 6$ and $N = 18$

	Treatment group		
	1	2	3
	4	2	17
	7	3	14
	10	11	12
	15	1	13
	9	8	16
	18	5	6
R_i	63	30	78

8.8.2 A Monte Carlo Analysis with $v = 2$

For the first analysis of the rank-score data listed in Table 8.10 under the Fisher–Pitman permutation model let $v = 2$, employing squared Euclidean scaling between the pairs of rank scores, and let the treatment-groups weights be given by

$$C_i = \frac{n_i - 1}{N - g}, \quad i = 1, \dots, g,$$

for correspondence with the Kruskal–Wallis g -sample rank-sum test. The average distance-function values for the $g = 3$ samples are

$$\xi_1 = 53.40, \quad \xi_2 = 29.60, \quad \text{and} \quad \xi_3 = 30.40,$$

and the observed value of the permutation test statistic based on $v = 2$ is

$$\delta = \sum_{i=1}^g C_i \xi_i = \frac{6 - 1}{18 - 3} (53.40 + 29.60 + 30.40) = 37.80.$$

Because there are

$$M = \frac{N!}{\prod_{i=1}^g n_i!} = \frac{18!}{6! 6! 6!} = 17,153,136$$

possible, equally-likely arrangements in the reference set of all permutations of the $N = 18$ rank scores listed in Table 8.10, an exact permutation analysis is not practical and a Monte Carlo permutation analysis is utilized.

Under the Fisher–Pitman permutation model, the Monte Carlo probability of an observed δ is the proportion of δ test statistic values computed on the randomly-selected, equally-likely arrangements of the $N = 18$ rank scores listed in Table 8.10 that are equal to or less than the observed value of $\delta = 37.80$. There are exactly 21,810 δ test statistic values that are equal to or less than the observed value of $\delta = 37.80$. If all M arrangements of the $N = 18$ observations listed in Table 8.10 occur with equal chance under the Fisher–Pitman null hypothesis, the Monte Carlo probability value of $\delta = 37.80$ computed on $L = 1,000,000$ random arrangements of the observed data with $n_1 = n_2 = n_3 = 6$ preserved for each arrangement is

$$P(\delta \leq \delta_o | H_0) = \frac{\text{number of } \delta \text{ values } \leq \delta_o}{L} = \frac{21,810}{1,000,000} = 0.0218,$$

where δ_o denotes the observed value of test statistic δ and L is the number of randomly-selected, equally-likely arrangements of the $N = 18$ rank scores listed in Table 8.10. It should be noted that whereas the Kruskal–Wallis test statistic, H , as

defined in Eq. (8.24) does not allow for tied rank scores, test statistic δ automatically accommodates tied rank scores.

The functional relationships between test statistics δ and H are given by

$$\delta = \frac{2 \left(T - \left\{ \frac{S}{6} [H + 3(N + 1)] \right\} \right)}{N - g} \quad (8.25)$$

and

$$H = \frac{6}{S} \left[T - \frac{\delta}{2} (N - g) \right] - 3(N + 1), \quad (8.26)$$

where, if no rank scores are tied, S and T may simply be expressed as

$$S = \sum_{i=1}^N i = \frac{N(N + 1)}{2} \quad \text{and} \quad T = \sum_{i=1}^N i^2 = \frac{N(N + 1)(2N + 1)}{6}.$$

Note that in Eqs. (8.25) and (8.26), S , T , N , and g are invariant under permutation, along with the constants 2, 3, and 6.

The relationships between test statistics δ and H can be confirmed with the rank-score data listed in Table 8.10. For the rank scores listed in Table 8.10 with no tied values, the observed value of S is

$$S = \sum_{i=1}^N i = \frac{N(N + 1)}{2} = \frac{18(18 + 1)}{2} = 171,$$

and the observed value of T is

$$T = \sum_{i=1}^N i^2 = \frac{N(N + 1)(2N + 1)}{6} = \frac{18(18 + 1)(2(18) + 1)}{6} = 2109.$$

Then following Eq. (8.25), the observed value of the permutation test statistic for the $N = 18$ rank scores listed in Table 8.10 is

$$\begin{aligned} \delta &= \frac{2 \left(T - \left\{ \frac{S}{6} [H + 3(N + 1)] \right\} \right)}{N - g} = \frac{N(N + 1)(N - 1 - H)}{6(N - g)} \\ &= \frac{18(18 + 1)(18 - 1 - 7.0526)}{6(18 - 3)} = 37.80 \end{aligned}$$

and, following Eq. (8.26), the observed value of the Kruskal–Wallis test statistic is

$$\begin{aligned} H &= \frac{6}{S} \left[T - \frac{\delta}{2}(N - g) \right] - 3(N + 1) = N - 1 - \frac{6\delta(N - g)}{N(N + 1)} \\ &= 18 - 1 - \frac{6(37.80)(18 - 3)}{18(18 + 1)} = 7.0526 . \end{aligned}$$

Because of the relationship between test statistics δ and H , the Monte Carlo probability value of the realized value of $H = 7.0526$ is identical to the Monte Carlo probability value of $\delta = 37.80$ under the Fisher–Pitman null hypothesis. Thus,

$$P(H \geq H_0 | H_0) = \frac{\text{number of } H \text{ values} \geq H_0}{L} = \frac{21,810}{1,000,000} = 0.0218 ,$$

where H_0 denotes the observed value of test statistic H .

The exact expected value of the $M = 17,153,136$ δ test statistic values under the Fisher–Pitman null hypothesis is

$$\mu_\delta = \frac{1}{M} \sum_{i=1}^M \delta_i = \frac{977,728,752}{17,153,136} = 57.00$$

and the observed chance-corrected measure of effect size is

$$\mathfrak{R} = 1 - \frac{\delta}{\mu_\delta} = 1 - \frac{37.80}{57.00} = +0.3368 ,$$

indicating approximately 34% within-group agreement above what is expected by chance. No comparisons are made with Cohen's \hat{d} , Pearson's η^2 (r^2), Kelley's $\hat{\eta}^2$, Hays' $\hat{\omega}^2$, or Hays' $\hat{\omega}_R^2$ measures of effect size as \hat{d} , η^2 , $\hat{\eta}^2$, $\hat{\omega}_F^2$, and $\hat{\omega}_R^2$ are undefined for rank-score data.

8.8.3 An Exact Analysis with $v = 2$

Although an exact permutation analysis with $M = 17,153,136$ possible arrangements of the observed data may be impractical, it is not impossible. An exact permutation analysis of the $N = 18$ observations listed in Table 8.10 yields $g = 3$ average distance-function values of

$$\xi_1 = 53.40 , \quad \xi_2 = 29.60 , \quad \text{and} \quad \xi_3 = 30.40 ,$$

and the observed value of the permutation test statistic based on $v = 2$ and treatment-group weights

$$C_i = \frac{N_i - 1}{N - g}, \quad i = 1, 2, 3,$$

is

$$\delta = \sum_{i=1}^g C_i \xi_i = \frac{6-1}{18-3} (53.40 + 29.60 + 30.40) = 37.80.$$

There are

$$M = \frac{N!}{\prod_{i=1}^g n_i!} = \frac{18!}{6! 6! 6!} = 17,153,136$$

possible, equally-likely arrangements in the reference set of all permutations of the $N = 18$ rank scores listed in Table 8.10, making an exact permutation analysis feasible. Under the Fisher–Pitman permutation model, the exact probability of an observed δ is the proportion of δ test statistic values computed on all possible, equally-likely arrangements of the $N = 18$ rank scores listed in Table 8.10 that are equal to or less than the observed value of $\delta = 37.80$. There are exactly 376,704 δ test statistic values that are equal to or less than the observed value of $\delta = 37.80$. If all M arrangements of the $N = 18$ rank scores listed in Table 8.10 occur with equal chance under the Fisher–Pitman null hypothesis, the exact probability value of $\delta = 37.80$ computed on the $M = 17,153,136$ possible arrangements of the observed data with $n_1 = n_2 = n_3 = 6$ preserved for each arrangement is

$$P(\delta \leq \delta_o | H_0) = \frac{\text{number of } \delta \text{ values } \leq \delta_o}{M} = \frac{376,704}{17,153,136} = 0.0220,$$

where δ_o denotes the observed value of test statistic δ and M is the number of possible, equally-likely arrangements of the $N = 18$ rank scores listed in Table 8.10. For comparison, the Monte Carlo probability value based on $v = 2$, $L = 1,000,000$ random arrangements of the observed data, and treatment-group weights given by

$$C_i = \frac{n_i - 1}{N - g}, \quad i = 1, 2, 3,$$

is $P = 0.0218$ for a difference between the two probability values of only

$$\Delta_P = 0.0220 - 0.0218 = 0.0002.$$

8.8.4 An Exact Analysis with $v = 1$

For a second analysis of the rank-score data listed in Table 8.10, let the treatment-group weights be given by

$$C_i = \frac{n_i - 1}{N - g}, \quad i = 1, \dots, g,$$

as in the previous example but set $v = 1$, employing ordinary Euclidean scaling between the pairs of rank scores. The $N = 18$ rank scores listed in Table 8.10 yield $g = 3$ average distance-function values of

$$\xi_1 = 6.3333, \quad \xi_2 = 4.6667, \quad \text{and} \quad \xi_3 = 4.5333,$$

and the observed value of the permutation test statistic based on $v = 1$ is

$$\delta = \sum_{i=1}^g C_i \xi_i = \frac{6 - 1}{18 - 3} (6.3333 + 4.6667 + 4.5333) = 5.1778.$$

Under the Fisher–Pitman permutation model, the exact probability of an observed δ is the proportion of δ test statistic values computed on all possible, equally-likely arrangements of the $N = 18$ rank scores listed in Table 8.10 that are equal to or less than the observed value of $\delta = 5.1778$. There are exactly 547,662 δ test statistic values that are equal to or less than the observed value of $\delta = 5.1778$. If all M arrangements of the $N = 18$ rank scores listed in Table 8.10 occur with equal chance under the Fisher–Pitman null hypothesis, the exact probability value of $\delta = 5.1778$ computed on the $M = 17,153,136$ possible arrangements of the observed data with $n_1 = n_2 = n_3 = 6$ preserved for each arrangement is

$$P(\delta \leq \delta_o | H_0) = \frac{\text{number of } \delta \text{ values } \leq \delta_o}{M} = \frac{547,662}{17,153,136} = 0.0319,$$

where δ_o denotes the observed value of test statistic δ and M is the number of possible, equally-likely arrangements of the $N = 18$ rank scores listed in Table 8.10. For comparison, the exact probability value based on $v = 2$, $M = 17,153,136$, and

$$C_i = \frac{n_i - 1}{N - g}, \quad i = 1, 2, 3,$$

is $P = 0.0220$. No comparison is made with the conventional Kruskal–Wallis g -sample rank-sum test as H is undefined for ordinary Euclidean scaling.

The exact expected value of the $M = 17,153,136$ δ test statistic values under the Fisher–Pitman null hypothesis is

$$\mu_\delta = \frac{1}{M} \sum_{i=1}^M \delta_i = \frac{108,636,5232}{17,153,136} = 6.3333,$$

and the observed chance-corrected measure of effect size is

$$\mathfrak{R} = 1 - \frac{\delta}{\mu_\delta} = 1 - \frac{5.1778}{6.3333} = +0.1825,$$

indicating approximately 18% within-group agreement above what is expected by chance. No comparisons are made with Cohen's \hat{d} , Pearson's r^2 (η^2), Kelley's $\hat{\eta}^2$, Hays' $\hat{\omega}_{\text{F}}^2$, or Hays' $\hat{\omega}_{\text{R}}^2$ measures of effect size as \hat{d} , r^2 , $\hat{\eta}^2$, $\hat{\omega}_{\text{F}}^2$, and $\hat{\omega}_{\text{R}}^2$ are undefined for rank-score data.

8.9 Example 6: Multivariate Permutation Analyses

It is sometimes desirable to test for differences among $g \geq 3$ independent treatment groups where $r \geq 2$ measurement scores have been obtained from each object. The conventional approach is a one-way multivariate analysis of variance (MANOVA) for which a number of statistical tests have been proposed, including the Bartlett–Nanda–Pillai (BNP) trace test [1, 16, 19], Wilks' likelihood-ratio test [26], Roy's maximum-root test [20, 21], and the Lawley–Hotelling trace test [9, 13, 14]. The Bartlett–Nanda–Pillai trace test is considered to be the most powerful and robust of the four tests [17, 18, 23, p. 269].

8.9.1 The Bartlett–Nanda–Pillai Trace Test

To illustrate a conventional multivariate analysis of variance, consider the BNP trace test given by

$$V^{(s)} = \text{trace} [\mathbf{B}(\mathbf{W} + \mathbf{B})^{-1}],$$

where \mathbf{W} denotes the Within matrix summarizing within-object variability, \mathbf{B} denotes the hypothesized Between matrix summarizing between-object variability, and $s = \min(r, g - 1)$. For a conventional test of significance, the BNP trace statistic, $V^{(s)}$, can be transformed into a conventional F test statistic by

$$F = \frac{2u + s + 1}{2t + s + 1} \left(\frac{V^{(s)}}{s - V^{(s)}} \right), \quad (8.27)$$

Table 8.11 Example multivariate response measurement scores with $r = 2$ measurement scores, $g = 3$ treatment groups, $n_1 = 5, n_2 = 4, n_3 = 3$, and $N = 12$ observations

Treatment group		
1	2	3
(5.8, 6.0)	(4.1, 2.9)	(4.2, 7.8)
(6.2, 3.9)	(3.9, 4.1)	(5.1, 5.9)
(3.9, 4.1)	(4.9, 3.9)	(4.8, 7.2)
(5.1, 5.2)	(2.1, 5.1)	
(3.0, 2.8)		

where $s = \min(r, g - 1)$, $u = 0.50(N - g - r - 1)$, $t = 0.50(|r - q| - 1)$, and $q = g - 1$. Assuming independence, normality, and homogeneity of variance and covariance, test statistic F is asymptotically distributed as Snedecor’s F under the Neyman–Pearson null hypothesis with $\nu_1 = s(2t + s + 1)$ and $\nu_2 = s(2u + s + 1)$ degrees of freedom.

To illustrate the BNP trace test, consider the multivariate observations listed in Table 8.11, where $r = 2$ measurements, $g = 3$ treatment groups, $n_1 = 5, n_2 = 4$, and $n_3 = 3$ sample sizes, and $N = 12$ multivariate observations.

A conventional BNP analysis of the multivariate observations listed in Table 8.11 yields

$$\mathbf{W} = \begin{bmatrix} 11.71000 & 1.17000 \\ 1.17000 & 10.42667 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 2.75250 & 3.19755 \\ 3.19755 & 17.30242 \end{bmatrix},$$

$$\mathbf{W} + \mathbf{B} = \begin{bmatrix} 14.46250 & 4.36755 \\ 4.36755 & 27.72909 \end{bmatrix},$$

$$(\mathbf{W} + \mathbf{B})^{-1} = \begin{bmatrix} 0.07260 & -0.01143 \\ -0.01143 & 0.03786 \end{bmatrix},$$

$$\mathbf{B}(\mathbf{W} + \mathbf{B})^{-1} = \begin{bmatrix} 0.16328 & 0.08960 \\ 0.03476 & 0.61852 \end{bmatrix},$$

and

$$V^{(2)} = \text{trace} [\mathbf{B}(\mathbf{W} + \mathbf{B})^{-1}] = 0.16328 + 0.61852 = 0.7818 .$$

Then, $q = g - 1 = 3 - 1 = 2$, $s = \min(r, q) = \min(2, 3 - 1) = 2$, $u = 0.50(N - g - r - 1) = 0.50(12 - 3 - 2 - 1) = 3$, $t = 0.50(|r - q| - 1) = 0.50(|2 - 2| - 1) = -0.50$, and following Eq. (8.27) on p. 306, the observed value

of Fisher's F -ratio test statistic is

$$F = \frac{2(3) + 2 + 1}{2(-0.50) + 2 + 1} \left(\frac{0.7818}{2 - 0.7818} \right) = \frac{9}{2}(0.6414) = 2.8879 .$$

Assuming independence, normality, homogeneity of variance, and homogeneity of covariance, test statistic F is asymptotically distributed as Snedecor's F with $v_1 = s(2t + s + 1) = 2[(2)(-0.50) + 2 + 1] = 4$ and $v_2 = s(2u + s + 1) = 2[(2)(3) + 2 + 1] = 18$ degrees of freedom. Under the Neyman–Pearson null hypothesis, the observed value of $F = 2.8879$ with $v_1 = 4$ and $v_2 = 18$ degrees of freedom yields an asymptotic probability value of $P = 0.0521$.

8.9.2 An Exact Analysis with $v = 2$

For the first analysis of the observed data listed in Table 8.11 under the Fisher–Pitman permutation model let $v = 2$, employing squared Euclidean scaling between the pairs of multivariate observations, and let the treatment-group weights be given by

$$C_i = \frac{n_i - 1}{N - g}, \quad i = 1, \dots, g ,$$

for correspondence with the BNP trace test. An exact permutation analysis is feasible for the multivariate observations listed in Table 8.11 as there are only

$$M = \frac{N!}{\prod_{i=1}^g n_i!} = \frac{12!}{5! 4! 3!} = 27,720$$

possible, equally-likely arrangements in the reference set of all permutations of the $N = 12$ multivariate scores listed in Table 8.11.

Following Eq. (8.2) on p. 261, the multivariate observations listed in Table 8.11 yield $g = 3$ average distance-function values of

$$\xi_1 = 0.3242, \quad \xi_2 = 0.2994, \quad \text{and} \quad \xi_3 = 0.1207 .$$

Following Eq. (8.1) on p. 261, the observed value of the permutation test statistic based on $v = 2$ and treatment-group weights

$$C_i = \frac{n_i - 1}{N - g}, \quad i = 1, 2, 3 ,$$

is

$$\delta = \sum_{i=1}^g C_i \xi_i = \frac{1}{12-3} [(5-1)(0.3242) + (4-1)(0.2994) + (3-1)(0.1207)] = 0.2707 .$$

Under the Fisher–Pitman permutation model, the exact probability of an observed δ is the proportion of δ test statistic values computed on all possible, equally-likely arrangements of the $N = 12$ multivariate observations listed in Table 8.11 that are equal to or less than the observed value of $\delta = 0.2707$. There are exactly 967 δ test statistic values that are equal to or less than the observed value of $\delta = 0.2707$. If all M arrangements of the $N = 12$ multivariate scores listed in Table 8.11 occur with equal chance under the Fisher–Pitman null hypothesis, the exact probability value of $\delta = 0.2707$ computed on the $M = 27,720$ possible arrangements of the observed data with $n_1 = 5$, $n_2 = 4$, and $n_3 = 3$ multivariate observations preserved for each arrangement is

$$P(\delta \leq \delta_o | H_0) = \frac{\text{number of } \delta \text{ values } \leq \delta_o}{M} = \frac{967}{27,720} = 0.0349 ,$$

where δ_o denotes the observed value of test statistic δ and M is the number of possible, equally-likely arrangements of the $N = 12$ multivariate observations listed in Table 8.11.

Following Eq. (8.7) on p. 263, the exact expected value of the $M = 27,720$ δ test statistic values under the Fisher–Pitman null hypothesis is

$$\mu_\delta = \frac{1}{M} \sum_{i=1}^M \delta_i = \frac{10,080}{27,720} = 0.3636$$

and, following Eq. (8.6) on p. 263, the observed chance-corrected measure of effect size is

$$\mathfrak{R} = 1 - \frac{\delta}{\mu_\delta} = 1 - \frac{0.2707}{0.3636} = +0.2556 ,$$

indicating approximately 26% within-group agreement above what is expected by chance.

A convenient, although positively biased, measure of effect size for the BNP trace test is given by

$$\eta^2 = \frac{V^{(2)}}{s} = \frac{0.7818}{2} = 0.3909 ,$$

which can be compared with the unbiased chance-corrected measure of effect size, $\mathfrak{R} = +0.2556$. No comparisons are made with Cohen's \hat{d} , Kelley's $\hat{\eta}^2$, Hays' $\hat{\omega}_F^2$, or Hays' $\hat{\omega}_R^2$ measures of effect size as \hat{d} , $\hat{\eta}^2$, $\hat{\omega}_F^2$, and $\hat{\omega}_R^2$ are undefined for multivariate data.

The functional relationships between statistic δ and the $V^{(2)}$ BNP trace statistic are given by

$$\delta = \frac{2(r - V^{(2)})}{N - g} \quad \text{and} \quad V^{(2)} = r - \frac{\delta(N - g)}{2}. \quad (8.28)$$

Following the expressions given in Eq. (8.28) for test statistics δ and $V^{(2)}$, the observed value for test statistic δ with respect to the observed value of test statistic $V^{(2)}$ is

$$\delta = \frac{2(r - V^{(2)})}{N - g} = \frac{2(2 - 0.7818)}{12 - 3} = 0.2707$$

and the observed value for test statistic $V^{(2)}$ with respect to the observed value of test statistic δ is

$$V^{(2)} = r - \frac{\delta(N - g)}{2} = 2 - \frac{(0.2707)(12 - 3)}{2} = 0.7818.$$

8.9.3 An Exact Analysis with $v = 1$

For a second analysis of the multivariate measurement scores listed in Table 8.11 on p. 307 under the Fisher–Pitman permutation model, let the treatment-group weights again be given by

$$C_i = \frac{n_i - 1}{N - g}, \quad i = 1, \dots, g,$$

but set $v = 1$ instead of $v = 2$, employing ordinary Euclidean scaling between the $N = 12$ multivariate scores. Following Eq. (8.2) on p. 261, the multivariate scores listed in Table 8.11 yield $g = 3$ average distance-function values of

$$\xi_1 = 2.3933, \quad \xi_2 = 1.9326, \quad \text{and} \quad \xi_3 = 1.4284.$$

Following Eq. (8.1) on p. 261, the observed value of the permutation test statistic based on $v = 1$ and treatment-group weights

$$C_i = \frac{n_i - 1}{N - g}, \quad i = 1, 2, 3,$$

is

$$\delta = \sum_{i=1}^g C_i \xi_i = \frac{1}{12-3} [(5-1)(2.3933) + (4-1)(1.9326) + (3-1)(1.4284)] = 2.0253 .$$

There are only

$$M = \frac{N!}{\prod_{i=1}^g n_i!} = \frac{12!}{5! 4! 3!} = 27,720$$

possible, equally-likely arrangements in the reference set of all permutations of the $N = 12$ multivariate observations listed in Table 8.11, making an exact permutation analysis feasible.

Under the Fisher–Pitman permutation model, the exact probability of an observed δ is the proportion of δ test statistic values computed on all possible, equally-likely arrangements of the $N = 12$ multivariate observations listed in Table 8.11 that are equal to or less than the observed value of $\delta = 2.0253$. There are exactly 618 δ test statistic values that are equal to or less than the observed value of $\delta = 2.0253$. If all M arrangements of the $N = 12$ multivariate observations listed in Table 8.11 occur with equal chance under the Fisher–Pitman null hypothesis, the exact probability value of $\delta = 2.0253$ computed on the $M = 27,720$ possible arrangements of the observed data with $n_1 = 5$, $n_2 = 4$, and $n_3 = 3$ multivariate observations preserved for each arrangement is

$$P(\delta \leq \delta_o | H_0) = \frac{\text{number of } \delta \text{ values } \leq \delta_o}{M} = \frac{618}{27,720} = 0.0223 ,$$

where δ_o denotes the observed value of test statistic δ and M is the number of possible, equally-likely arrangements of the $N = 12$ multivariate observations listed in Table 8.11. No comparison is made with the Bartlett–Nanda–Pillai trace test as the BNP test is undefined for ordinary Euclidean scaling.

Following Eq. (8.7) on p. 263, the exact expected value of the $M = 27,720$ δ test statistic values under the Fisher–Pitman null hypothesis is

$$\mu_\delta = \frac{1}{M} \sum_{i=1}^M \delta_i = \frac{69,854}{27,720} = 2.5200$$

and, following Eq. (8.6) on p. 263, the observed chance-corrected measure of effect size is

$$\mathfrak{R} = 1 - \frac{\delta}{\mu_\delta} = 1 - \frac{2.0253}{2.5200} = +0.1963 ,$$

indicating approximately 20% within-group agreement above that expected by chance. No comparison is made with the conventional measure of effect size as η^2 is undefined for ordinary Euclidean scaling.

8.10 Summary

This chapter examined statistical methods for multiple independent samples where the null hypothesis posits no differences among the $g \geq 3$ populations that the g random samples are presumed to represent. Under the Neyman–Pearson population model of statistical inference, a conventional one-way analysis of variance and five measures of effect size were described and illustrated: Fisher’s F -ratio test statistic, and Cohen’s \hat{d} , Pearson’s η^2 , Kelley’s $\hat{\eta}^2$, Hays’ $\hat{\omega}_F^2$, and Hays’ $\hat{\omega}_R^2$ measures of effect size, respectively.

Under the Fisher–Pitman permutation model of statistical inference, test statistic δ and associated measure of effect size, \mathfrak{R} , were described and illustrated for multi-sample tests. For tests of $g \geq 3$ independent samples, test statistic δ was demonstrated to be flexible enough to incorporate both ordinary and squared Euclidean scaling functions with $v = 1$ and $v = 2$, respectively. Effect size measure \mathfrak{R} was shown to be applicable to either $v = 1$ or $v = 2$ without modification and to have a clear and meaningful chance-corrected interpretation.

Six examples illustrated permutation-based statistics δ and \mathfrak{R} . In the first example, a small sample of $N = 10$ observations in $g = 3$ treatment groups was utilized to describe and illustrate the calculation of test statistics δ and \mathfrak{R} for multiple independent samples. The second example with $N = 10$ observations in $g = 3$ treatment groups demonstrated the chance-corrected measure of effect size, \mathfrak{R} , and related \mathfrak{R} to the five conventional measures of effect size for $g \geq 3$ independent samples: Cohen’s \hat{d} , Pearson’s η^2 , Kelley’s $\hat{\eta}^2$, Hays’ $\hat{\omega}_F^2$, and Hays’ $\hat{\omega}_R^2$. The third example with $N = 28$ observations in $g = 4$ treatment groups illustrated the effects of extreme values on analyses using $v = 1$ for ordinary Euclidean scaling and $v = 2$ for squared Euclidean scaling. The fourth example with $N = 15$ observations in $g = 4$ treatment groups compared exact and Monte Carlo permutation statistical methods, illustrating the accuracy and efficiency of Monte Carlo analyses. The fifth example with $N = 18$ rank scores in $g = 3$ treatment groups illustrated an application of permutation statistical methods to univariate rank-score data, comparing a permutation analysis of the rank-score data with the conventional Kruskal–Wallis g -sample one-way analysis of variance for ranks. In the sixth example, both test statistic δ and effect size measure \mathfrak{R} were extended to multivariate data with $N = 12$ multivariate observations in $g = 3$ treatment groups and compared the permutation analysis of the multivariate data to the conventional Bartlett–Nanda–Pillai trace test for multivariate independent samples.

Chapter 9 continues the presentation of permutation statistical methods for $g \geq 3$ samples, but examines research designs in which the subjects in the $g \geq 3$ samples are matched on specific characteristics; that is, not independent. Research designs

that posit no differences among matched treatment groups have a long history and are ubiquitous in the contemporary statistical literature and are generally known as randomized-blocks designs, of which there exist a large variety.

References

1. Bartlett, M.S.: A note on tests of significance in multivariate analysis. *Proc. Camb. Philos. Soc.* **34**, 33–40 (1939)
2. Berry, K.J., Mielke, P.W., Johnston, J.E.: *Permutation Statistical Methods: An Integrated Approach*. Springer, Cham (2016)
3. Boik, R.J.: The Fisher–Pitman permutation test: a non-robust alternative to the normal theory F test when variances are heterogeneous. *Brit. J. Math. Stat. Psychol.* **40**, 26–42 (1987)
4. Box, J.F.: *R. A. Fisher: The Life of a Scientist*. Wiley, New York (1978)
5. Festinger, L.: The significance of differences between means without reference to the frequency distribution function. *Psychometrika* **11**, 97–105 (1946)
6. Haggard, E.A.: *Intraclass Correlation and the Analysis of Variance*. Dryden, New York (1958)
7. Haldane, J.B.S., Smith, C.A.B.: A simple exact test for birth-order effect. *Ann. Eugenetic.* **14**, 117–124 (1948)
8. Hall, N.S.: R. A. Fisher and his advocacy of randomization. *J. Hist. Biol.* **40**, 295–325 (2007)
9. Hotelling, H.: A generalized T test and measure of multivariate dispersion. In: Neyman, J. (ed.) *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, vol. II, pp. 23–41. University of California Press, Berkeley (1951)
10. Hotelling, H., Pabst, M.R.: Rank correlation and tests of significance involving no assumption of normality. *Ann. Math. Stat.* **7**, 29–43 (1936)
11. Johnston, J.E., Berry, K.J., Mielke, P.W.: Permutation tests: precision in estimating probability values. *Percept. Motor Skill.* **105**, 915–920 (2007)
12. 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)]
13. Lawley, D.N.: A generalization of Fisher’s z test. *Biometrika* **30**, 180–187 (1938)
14. Lawley, D.N.: Corrections to “A generalization of Fisher’s z test”. *Biometrika* **30**, 467–469 (1939)
15. 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)
16. Nanda, D.N.: Distribution of the sum of roots of a determinantal equation. *Ann. Math. Stat.* **21**, 432–439 (1950)
17. Olson, C.L.: On choosing a test statistic in multivariate analysis of variance. *Psychol. Bull.* **83**, 579–586 (1976)
18. Olson, C.L.: Practical considerations in choosing a MANOVA test statistic: a rejoinder to Stevens. *Psychol. Bull.* **86**, 1350–1352 (1979)
19. Pillai, K.C.S.: Some new test criteria in multivariate analysis. *Ann. Math. Stat.* **26**, 117–121 (1955)
20. Roy, S.N.: On a heuristic method of test construction and its use in multivariate analysis. *Ann. Math. Stat.* **24**, 220–238 (1953)
21. Roy, S.N.: *Some Aspects of Multivariate Analysis*. Wiley, New York (1957)
22. Snedecor, G.W.: *Calculation and Interpretation of Analysis of Variance and Covariance*. Collegiate Press, Ames (1934)
23. Tabachnick, B.G., Fidell, L.S.: *Using Multivariate Statistics*, 5th edn. Pearson, Boston (2007)
24. van der Reyden, D.: A simple statistical significance test. *Rhod. Agric. J.* **49**, 96–104 (1952)
25. Wilcoxon, F.: Individual comparisons by ranking methods. *Biometrics Bull.* **1**, 80–83 (1945)
26. Wilks, S.S.: Certain generalizations in the analysis of variance. *Biometrika* **24**, 471–494 (1932)