Chapter 9
Randomized-Blocks Designs

Abstract This chapter introduces permutation methods for multiple matched samples, i.e., randomized-blocks designs. Included in this chapter are six example analyses illustrating computation of exact permutation probability values for randomized-blocks designs, calculation of measures of effect size for randomized-blocks designs, the effect of extreme values on conventional and permutation randomized-blocks designs, exact and Monte Carlo permutation procedures for randomized-blocks designs, application of permutation methods to randomized-blocks designs with rank-score data, and analysis of randomized-blocks designs with multivariate data. Included in this chapter are permutation versions of Fisher's F test for a one-way randomized-blocks design, Friedman's two-way analysis of variance for ranks, and a permutation-based alternative for the four conventional measures of effect size for randomized-blocks designs: Hays' $\hat{\omega}^2$, Pearson's η^2, Cohen's partial η^2, and Cohen's f^2.

This chapter presents exact and Monte Carlo permutation statistical methods for tests of experimental differences among three or more matched or otherwise related samples, commonly called randomized-blocks designs under the Neyman–Pearson population model of statistical inference. As with matched-pairs tests discussed in Chap. 7, the samples may either be matched on specific criteria; for example, age, education, gender, or the same subjects may be observed at different times or under different treatments or interventions. While most randomized-blocks designs take observations at two, three, or four time periods, there have been a number of long-running studies that follow clients over many years. The best-known of these are the Fels Longitudinal Study founded in 1929 as a division of the Fels Research Institute in Yellow Springs, Ohio, the Framingham Heart Study initiated in 1948 in Framingham, Massachusetts, and the Terman Genetic Study of Genius founded at Stanford University in 1921. All three studies continue today.[1]

[1] Studies such as these that observe the same or matched subjects for many years are often referred to as "panel studies" and require a different statistical approach.

K. J. Berry et al., *A Primer of Permutation Statistical Methods*,
https://doi.org/10.1007/978-3-030-20933-9_9

As in previous chapters, six examples illustrate permutation statistical methods for randomized-blocks designs. The first example utilizes a small set of data to illustrate the computation of exact permutation methods for multiple matched samples, wherein the permutation test statistic, δ, is developed and compared with Fisher's conventional F-ratio test statistic for multiple dependent samples. The second example develops a permutation-based measure of effect size as a chance-corrected alternative to the four conventional measures of effect size for randomized-blocks designs: Hays' $\hat{\omega}^2$, Pearson's η^2, Cohen's partial η^2, and Cohen's f^2. The third example compares permutation statistical methods based on ordinary and squared Euclidean scaling functions, with an emphasis on the analysis of data sets containing extreme values. The fourth example utilizes a larger set of data to provide a comparison of exact permutation methods and Monte Carlo permutation methods, demonstrating the efficiency and accuracy of Monte Carlo permutation statistical methods for multiple matched samples. The fifth example illustrates the application of permutation statistical methods to univariate rank-score data, comparing permutation statistical methods to Friedman's conventional two-way analysis of variance for ranks. The sixth example illustrates the application of permutation statistical methods to multivariate data.

9.1 Introduction

The standard univariate test for $g \geq 3$ matched samples under the Neyman–Pearson population model of inference is Fisher's randomized-blocks analysis of variance wherein the null hypothesis (H_0) posits no mean differences among the g populations from which the samples presumably have been randomly drawn; that is, H_0: $\mu_1 = \mu_2 = \cdots = \mu_g$. Fisher's randomized-blocks analysis of variance 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 randomly drawn from populations with identical mean values, assuming normality.

Consider samples of $N = bg$ independent random variables x_{ij} with cumulative distribution functions $F_i(x + \beta_j)$ for $i = 1, \ldots, g$ and $j = 1, \ldots, b$, respectively, where g denotes the number of treatments and b denotes the number of blocks. For simplicity, assume that the x_{ij} values are randomly drawn from a normal distribution with mean $\mu_i + \beta_j$ and variance σ_x^2, $i = 1, \ldots, g$ and $j = 1, \ldots, b$. Under the Neyman–Pearson population model, the null hypothesis of no mean differences tests

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

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 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 assumption of normality.

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

$$F = \frac{MS_{\text{Treatments}}}{MS_{\text{Error}}} ,$$

where the mean-square treatments is given by

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

the sum-of-squares treatments is given by

$$SS_{\text{Treatments}} = b \sum_{i=1}^{g} \left(\bar{x}_{i.} - \bar{x}_{..}\right)^2 ,$$

the mean-square error is given by

$$MS_{\text{Error}} = \frac{SS_{\text{Error}}}{(b-1)(g-1)} ,$$

the sum-of-squares error is given by

$$SS_{\text{Error}} = \sum_{i=1}^{g} \sum_{j=1}^{b} \left(x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..}\right)^2 ,$$

the sum-of-squares blocks is given by

$$SS_{\text{Blocks}} = g \sum_{j=1}^{b} \left(\bar{x}_{.j} - \bar{x}_{..}\right)^2 ,$$

the sum-of-squares total is given by

$$SS_{\text{Total}} = \sum_{i=1}^{g} \sum_{j=1}^{b} \left(x_{ij} - \bar{x}_{..}\right)^2 ,$$

the mean value for the ith of g treatments is

$$\bar{x}_{i.} = \frac{1}{b} \sum_{j=1}^{b} x_{ij} , \qquad i = 1, \ldots, g ,$$

the mean value for the jth of b blocks is

$$\bar{x}_{\cdot j} = \frac{1}{g}\sum_{i=1}^{g} x_{ij} \ , \qquad j = 1, \ \ldots, \ b \ ,$$

the grand mean over all b blocks and g treatments is given by

$$\bar{x}_{\cdot\cdot} = \frac{1}{gb}\sum_{i=1}^{g}\sum_{j=1}^{b} x_{ij} \ ,$$

and x_{ij} denotes the value of the jth block in the ith treatment for $j = 1, \ \ldots, \ b$ and $i = 1, \ \ldots, \ g$.

Under the Neyman–Pearson null hypothesis, H_0: $\mu_1 = \mu_2 = \cdots = \mu_g$, Fisher's F-ratio test statistic is asymptotically distributed as Snedecor's F with $\nu_1 = g - 1$ degrees of freedom (df) in the numerator and $\nu_2 = (b - 1)(g - 1)$ df in the denominator. If the x_{ij} values, $i = 1, \ \ldots, \ g$ and $j = 1, \ \ldots, \ b$, are not randomly sampled from a normally-distributed population, then Fisher's F-ratio test statistic no longer follows Snedecor's F distribution with $\nu_1 = g-1$ and $\nu_2 = (b-1)(g-1)$ degrees of freedom.

The assumptions underlying Fisher's F test for multiple matched samples are (1) the observations are independent, (2) the data are random samples from well-defined, normally-distributed populations, (3) homogeneity of variance, and (4) homogeneity of covariance.

9.2 A Permutation Approach

Alternatively, consider a test for multiple matched 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 [4]. Moreover, there is no alternative hypothesis under the permutation model and no specified α level. Also, there is no requirement of random sampling, no degrees of freedom, no assumption of normality, no assumption of homogeneity of variance, and no assumption of homogeneity of covariance. This is not to imply that the results of permutation statistical methods are unaffected by homogeneity of variance and covariance, but homogeneity of variance and covariance are not requirements for permutation methods as they are for conventional statistical methods under the Neyman–Pearson population model.

A permutation alternative to Fisher's conventional F-ratio test for $g \geq 3$ matched samples is given by

$$\delta = \left[g\binom{b}{2}\right]^{-1} \sum_{i=1}^{g}\sum_{j=1}^{b-1}\sum_{k=j+1}^{b} \Delta\left(x_{ij}, x_{ik}\right) \ , \tag{9.1}$$

where the symmetric distance functions are given by

$$\Delta(x, y) = \left[\left(x_i - y_i\right)^2\right]^{v/2} \tag{9.2}$$

and $v > 0$. When $v = 1$, ordinary Euclidean scaling is employed, and when $v = 2$, squared Euclidean scaling is employed [7].

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

$$M = \left(g!\right)^b$$

possible allocations of the observations to the g treatments within each of the b blocks. The probability value associated with an observed value of δ is the probability under the Fisher–Pitman null hypothesis of observing a value of δ that is equal to or less than the observed value of δ. An exact probability value for δ may be expressed as

$$P(\delta \leq \delta_{\mathrm{o}} | H_0) = \frac{\text{number of } \delta \text{ values } \leq \delta_{\mathrm{o}}}{M},$$

where δ_{o} denotes the observed value of test statistic δ and M is the number of possible, equally-likely arrangements of the observed data.

When M is large, an approximate probability value for test statistic δ may be obtained from a Monte Carlo procedure, where a large random sample of arrangements of the observed data is drawn. Then an approximate probability value for test statistic δ is given by

$$P(\delta \leq \delta_{\mathrm{o}} | H_0) = \frac{\text{number of } \delta \text{ values } \leq \delta_{\mathrm{o}}}{L},$$

where L denotes the number of the randomly-selected, equally-likely arrangements of the observed data.

9.3 The Relationship Between Statistics F and δ

When the null hypothesis under the Neyman–Pearson population model states H_0: $\mu_1 = \mu_2 = \cdots = \mu_g$ and $v = 2$, the functional relationships between test statistic δ and Fisher's F test statistic are given by

$$F = \frac{(b-1)[2SS_{\mathrm{Total}} - g(b-1)\delta]}{g(b-1)\delta - 2SS_{\mathrm{Blocks}}} \tag{9.3}$$

and

$$\delta = \frac{2[FSS_{\text{Blocks}} + (b-1)SS_{\text{Total}}]}{g(b-1)(F+b-1)} \,. \tag{9.4}$$

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

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

and

$$P(F \geq F_{\text{o}}|H_0) = \frac{\text{number of } F \text{ values } \geq F_{\text{o}}}{M} \,.$$

are equivalent under the Fisher–Pitman null hypothesis, where δ_{o} and F_{o} denote the observed values of test statistics δ and F, respectively, and M is the number of possible, equally-likely arrangements of the observed data.

A chance-corrected measure of agreement is given by

$$\Re = 1 - \frac{\delta}{\mu_\delta} \,, \tag{9.5}$$

where μ_δ, the exact expected value of the M δ test statistic values calculated on all possible arrangements of the observed measurements, is given by

$$\mu_\delta = \frac{1}{M}\sum_{i=1}^{M}\delta_i \,. \tag{9.6}$$

9.4 Example 1: Test Statistics F and δ

A small example will serve to illustrate the relationships between test statistics F and δ. Consider the example randomized-blocks data listed in Table 9.1 with $g = 2$ treatment groups, $b = 4$ blocks, and $N = bg = (4)(2) = 8$ total observations.

Table 9.1 Example data with $g = 2$ treatments and $b = 4$ blocks

	Treatment	
Block	1	2
1	105	21
2	144	52
3	109	97
4	113	32

Under the Neyman–Pearson population model with treatment means $\bar{x}_{1.} = 117.75$ and $\bar{x}_{2.} = 50.50$, block means $\bar{x}_{.1} = 63.00$, $\bar{x}_{.2} = 98.00$, $\bar{x}_{.3} = 103.00$, and $\bar{x}_{.4} = 72.50$, grand mean $\bar{x}_{..} = 84.1250$, the sum-of-squares total is

$$SS_{\text{Total}} = \sum_{i=1}^{g} \sum_{j=1}^{b} \left(x_{ij} - \bar{x}_{..}\right)^2 = 13{,}372.8750 ,$$

the sum-of-squares treatments is

$$SS_{\text{Treatments}} = b \sum_{i=1}^{g} \left(\bar{x}_{i.} - \bar{x}_{..}\right)^2 = 9045.1250 ,$$

the mean-square treatments is

$$MS_{\text{Treatments}} = \frac{SS_{\text{Treatments}}}{g-1} = \frac{9045.1250}{2-1} = 9045.1250 ,$$

the sum-of-squares blocks is

$$SS_{\text{Blocks}} = g \sum_{j=1}^{b} \left(\bar{x}_{.j} - \bar{x}_{..}\right)^2 = 2260.3750 ,$$

the sum-of-squares error is

$$SS_{\text{Error}} = \sum_{i=1}^{g} \sum_{j=1}^{b} \left(x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..}\right)^2 = 2067.3750 ,$$

the mean-square error is

$$MS_{\text{Error}} = \frac{SS_{\text{Error}}}{(b-1)(g-1)} = \frac{2067.3750}{(4-1)(2-1)} = 689.1250 ,$$

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

$$F = \frac{MS_{\text{Treatments}}}{MS_{\text{Error}}} = \frac{9045.1250}{689.1250} = 13.1255 .$$

For computational efficiency, SS_{Error} can easily be obtained by simple subtraction; for example,

$$\begin{aligned} SS_{\text{Error}} &= SS_{\text{Total}} - SS_{\text{Blocks}} - SS_{\text{Treatments}} \\ &= 13{,}372.8750 - 2260.3750 - 9045.1250 = 2067.3750 . \end{aligned}$$

Table 9.2 Source table for the data listed in Table 9.1

Factor	*SS*	*df*	*MS*	*F*
Blocks	2260.3750			
Treatments	9045.1250	1	9045.1250	13.1255
Error	2067.3750	3	689.1250	
Total	13,372.8750			

The essential factors, sums of squares (*SS*), degrees of freedom (*df*), mean squares (*MS*), and variance-ratio test statistic (F) are summarized in Table 9.2.

Under the Neyman–Pearson null hypothesis, H_0: $\mu_1 = \mu_2 = \cdots = \mu_g$, Fisher's F-ratio test statistic is asymptotically distributed as Snedecor's F with $\nu_1 = g - 1$ and $\nu_2 = (b-1)(g-1)$ degrees of freedom. With $\nu_1 = g-1 = 2-1 = 1$ and $\nu_2 = (b-1)(g-1) = (4-1)(2-1) = 3$ degrees of freedom, the asymptotic probability value of $F = 13.1255$ is $P = 0.0362$, under the assumptions of normality and homogeneity.

9.4.1 An Exact Analysis with $v = 2$

For an exact analysis under the Fisher–Pitman permutation model let $v = 2$, employing squared Euclidean scaling for correspondence with Fisher's F-ratio test statistic. Following Eq. (9.2) on p. 319 with $v = 2$ for Treatment 1, the six distance-function values are

$$\Delta(1,2) = \left(|105 - 144|^2\right)^{2/2} = 1521\,,$$

$$\Delta(1,3) = \left(|105 - 109|^2\right)^{2/2} = \quad 16\,,$$

$$\Delta(1,4) = \left(|105 - 113|^2\right)^{2/2} = \quad 64\,,$$

$$\Delta(2,3) = \left(|144 - 109|^2\right)^{2/2} = 1225\,,$$

$$\Delta(2,4) = \left(|144 - 113|^2\right)^{2/2} = \quad 961\,,$$

$$\Delta(3,4) = \left(|109 - 113|^2\right)^{2/2} = \quad 16\,,$$

and for Treatment 2, the six distance-function values are

$$\Delta(1,2) = \left(|21-52|^2\right)^{2/2} = 961\,,$$
$$\Delta(1,3) = \left(|21-97|^2\right)^{2/2} = 5776\,,$$
$$\Delta(1,4) = \left(|21-32|^2\right)^{2/2} = 121\,,$$
$$\Delta(2,3) = \left(|52-97|^2\right)^{2/2} = 2025\,,$$
$$\Delta(2,4) = \left(|52-32|^2\right)^{2/2} = 400\,,$$
$$\Delta(3,4) = \left(|97-32|^2\right)^{2/2} = 4225\,.$$

Following Eq. (9.1) on p. 318, the observed value of test statistic δ is

$$\begin{aligned}\delta &= \left[g\binom{b}{2}\right]^{-1} \sum_{i=1}^{g} \sum_{j=1}^{b-1} \sum_{k=j+1}^{b} \Delta\left(x_{ij}, x_{ik}\right) \\ &= \left[2\binom{4}{2}\right]^{-1} \left[\Delta(1,2) + \Delta(1,3) + \cdots + \Delta(3,4)\right] \\ &= \frac{1}{12}\left(1521 + 16 + 64 + \cdots + 400 + 4225\right) = 1442.5833\,.\end{aligned}$$

Alternatively, in terms of a randomized-blocks analysis of variance model the observed permutation test statistic is

$$\begin{aligned}\delta &= \frac{2(SS_{\text{Total}} - SS_{\text{Treatments}})}{N-g} \\ &= \frac{2(13{,}372.8750 - 9045.1250)}{8-2} = 1442.5833\,.\end{aligned}$$

Based on the expressions given in Eqs. (9.3) and (9.4) on p. 319, the observed value of test statistic F with respect to the observed value of test statistic δ is

$$\begin{aligned}F &= \frac{(b-1)[2SS_{\text{Total}} - g(b-1)\delta]}{g(b-1)\delta - 2SS_{\text{Blocks}}} \\ &= \frac{(4-1)[2(13{,}372.8750) - (2)(4-1)(1442.5833)]}{2(4-1)(1442.5833) - 2(2260.3750)} = 13.1255\end{aligned}$$

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

$$\delta = \frac{2[FSS_{\text{Blocks}} + (b-1)SS_{\text{Total}}]}{g(b-1)(F+b-1)}$$
$$= \frac{2[(13.1255)(2260.3750) + (4-1)(13{,}372.8750)]}{2(4-1)(13.1255+4-1)} = 1442.5833\,.$$

Because there are only

$$M = (g!)^b = (2!)^4 = 16$$

possible, equally-likely arrangements in the reference set of all permutations of the $N = 8$ observations listed in Table 9.1, an exact permutation analysis is 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 $N = 8$ observations listed in Table 9.1 that are equal to or less than the observed value of $\delta = 1442.5833$. Table 9.3 lists the $M = 16$ possible δ values, ordered from the lowest ($\delta_1 = 1442.5833$) to the highest ($\delta_{16} = 4302.5833$).

It is readily apparent from Table 9.3 that there are duplicate arrangements of the observed scores and duplicate δ values; for example, Order 1 {105, 144, 109, 113} minus Order 2 {21, 52, 97, 32} yields the same absolute difference as Order 15 {21, 52, 97, 32} minus Order 16 {105, 144, 109, 113}. It is more efficient to fix the

Table 9.3 Permutations of the observed scores listed in Table 9.1 with values for δ based on $v = 2$ ordered from lowest to highest

Order	Treatment 1	Treatment 2	δ
1	{105, 144, 109, 113}	{ 21, 52, 97, 32}	1442.5833
2	{ 21, 52, 97, 32}	{105, 144, 109, 113}	1442.5833
3	{105, 144, 97, 113}	{ 21, 52, 109, 32}	1956.5833
4	{ 21, 52, 109, 32}	{105, 144, 97, 113}	1956.5833
5	{ 21, 52, 97, 113}	{105, 144, 109, 32}	3980.5833
6	{105, 144, 109, 32}	{ 21, 52, 97, 113}	3980.5833
7	{ 21, 144, 109, 113}	{105, 52, 97, 32}	4032.5833
8	{105, 52, 97, 32}	{ 21, 144, 109, 113}	4032.5833
9	{105, 52, 109, 113}	{ 21, 144, 97, 32}	4156.5833
10	{ 21, 144, 97, 32}	{105, 52, 109, 113}	4156.5833
11	{ 21, 52, 109, 113}	{105, 144, 97, 32}	4170.5833
12	{105, 144, 97, 32}	{ 21, 52, 109, 113}	4170.5833
13	{ 21, 144, 97, 113}	{105, 52, 109, 32}	4210.5833
14	{105, 52, 109, 32}	{ 21, 144, 97, 113}	4210.5833
15	{105, 52, 97, 113}	{ 21, 144, 109, 32}	4302.5833
16	{ 21, 144, 109, 32}	{105, 52, 97, 113}	4302.5833

Table 9.4 Permutations of the observed scores listed in Table 9.1 with values for δ based on $v = 2$ ordered from lowest to highest

Order	Treatment 1	Treatment 2	δ
1	{105, 144, 109, 113}	{ 21, 52, 97, 32}	1442.5833
2	{105, 144, 97, 113}	{ 21, 52, 109, 32}	1956.5833
3	{ 21, 52, 97, 113}	{105, 144, 109, 32}	3980.5833
4	{ 21, 144, 109, 113}	{105, 52, 97, 32}	4032.5833
5	{105, 52, 109, 113}	{ 21, 144, 97, 32}	4156.5833
6	{ 21, 52, 109, 113}	{105, 144, 97, 32}	4170.5833
7	{ 21, 144, 97, 113}	{105, 52, 109, 32}	4210.5833
8	{105, 52, 97, 113}	{ 21, 144, 109, 32}	4302.5833
Total			28,252.6667

scores in one block and permute the remaining blocks. Thus,

$$M = (g!)^b = (2!)^4 = 16 \quad \text{is replaced by} \quad M = (g!)^{b-1} = (2!)^{4-1} = 8$$

and the results are listed in Table 9.4. There is only one δ value in Table 9.4 that is equal to or less than the observed value of $\delta = 1442.5833$. If all M arrangements of the $N = 8$ observations listed in Table 9.4 occur with equal chance under the Fisher–Pitman null hypothesis, the exact probability value of $\delta = 1442.5833$ computed on all $M = 8$ arrangements of the observed data with $b = 4$ blocks preserved for each arrangement is

$$P(\delta \leq \delta_o | H_0) = \frac{\text{number of } \delta \text{ values } \leq \delta_o}{M} = \frac{1}{8} = 0.1250\,,$$

where δ_o denotes the observed value of test statistic δ and M is the number of possible, equally-likely arrangements of the $N = 8$ observations listed in Table 9.1.

Alternatively, there is only one F value that is equal to or greater than the observed value of $F = 13.1255$, as illustrated in Table 9.5. Thus if all M arrangements of the $N = 8$ observations listed in Table 9.1 occur with equal

Table 9.5 Permutations of the observed scores listed in Table 9.1 with values for Fisher's F-ratio ordered from highest to lowest

Order	Treatment 1	Treatment 2	F-ratio
1	{105, 144, 109, 113}	{ 21, 52, 97, 32}	13.1255
2	{105, 144, 97, 113}	{ 21, 52, 109, 32}	6.2364
3	{ 21, 52, 97, 113}	{105, 144, 109, 32}	0.4435
4	{ 21, 144, 109, 113}	{105, 52, 97, 32}	0.3889
5	{105, 52, 109, 113}	{ 21, 144, 97, 32}	0.2654
6	{ 21, 52, 109, 113}	{105, 144, 97, 32}	0.2520
7	{ 21, 144, 97, 113}	{105, 52, 109, 32}	0.2144
8	{105, 52, 97, 113}	{ 21, 144, 109, 32}	0.1311

chance under the Fisher–Pitman null hypothesis, the exact probability value of $F = 13.1255$ is

$$P(F \geq F_o | H_0) = \frac{\text{number of } F \text{ values } \geq F_o}{M} = \frac{1}{8} = 0.1250 ,$$

where F_o denotes the observed value of test statistic F.

There is a considerable difference between the conventional asymptotic probability value for F ($P = 0.0362$) and the exact permutation probability value for δ ($P = 0.1250$). The difference between the two probability values of

$$\Delta_P = 0.1250 - 0.0362 = 0.0888$$

is most likely due to the very small number of blocks. A continuous mathematical function such as Snedecor's F cannot be expected to provide a precise fit to only $M = 8$ discrete data points.

Following Eq. (9.6) on p. 320, the exact expected value of the $M = 8$ δ test statistic values under the Fisher–Pitman null hypothesis is

$$\mu_\delta = \frac{1}{M} \sum_{i=1}^{M} \delta_i = \frac{28{,}252.6667}{8} = 3531.5833 .$$

Following Eq. (9.5) on p. 320, the observed chance-corrected measure of effect size is

$$\Re = 1 - \frac{\delta}{\mu_\delta} = 1 - \frac{1442.5833}{3531.5833} = +0.5915 ,$$

indicating approximately 59% within-block agreement above what is expected by chance.

9.5 Example 2: Measures of Effect Size

Many researchers deplore the sole reliance on tests of statistical significance and recommend that indices of effect size—magnitude of experimental effects—accompany tests of significance. Measures of effect size express the practical or clinical significance of differences among sample means, as contrasted with the statistical significance of the differences. Consequently, the reporting of measures of effect size in addition to tests of significance has become increasingly important in the contemporary research literature. For example, a 2018 article in *The Lancet* sought to establish the risk thresholds for alcohol consumption using a meta-analysis for 83 observational studies with a total of 599,912 consumers of alcohol,

concluding that no level of alcohol consumption is safe [11]. A critique of the article in the *New York Times* noted that no measure of effect size was included:

> [W]hen we compile observational study on top of observational study, we become more likely to achieve statistical significance without improving clinical significance. In other words, very small differences are real, but that doesn't mean those differences are critical [1, p. A12].

Five conventional measures of effect size for randomized-blocks analysis of variance designs are described and compared in this section: Hays' $\hat{\omega}^2$, Pearson's η^2, Cohen's partial η^2, Cohen's f^2, and Mielke and Berry's $\Re$.

Hays' $\hat{\omega}^2$ measure of effect size is given by

$$\hat{\omega}^2 = \frac{(g-1)(MS_{\text{Treatments}} - MS_{\text{Error}})}{SS_{\text{Total}} + MS_{\text{Within Blocks}}} , \tag{9.7}$$

where the mean-square within blocks is given by

$$MS_{\text{Within Blocks}} = \frac{SS_{\text{Within Blocks}}}{b(g-1)} ,$$

b and g denote the number of blocks and treatments, respectively, and the sum-of-squares within blocks is given by

$$SS_{\text{Within Blocks}} = SS_{\text{Total}} - SS_{\text{Blocks}} . \tag{9.8}$$

Pearson's η^2 measure of effect size is given by[2]

$$\eta^2 = \frac{SS_{\text{Treatments}}}{SS_{\text{Total}}} . \tag{9.9}$$

Cohen's partial η^2 measure of effect size is given by

$$\eta^2_{\text{Partial}} = \frac{SS_{\text{Treatments}}}{SS_{\text{Total}} - SS_{\text{Error}}} . \tag{9.10}$$

Cohen's f^2 measure of effect size is given by

$$f^2 = \frac{SS_{\text{Treatments}}}{SS_{\text{Total}} - SS_{\text{Treatments}}} . \tag{9.11}$$

Mielke and Berry's chance-corrected measure of effect size is given by

$$\Re = 1 - \frac{\delta}{\mu_\delta} , \tag{9.12}$$

[2]Pearson's η^2 measure of effect size is often erroneously referred to as the "correlation ratio." Technically, η is the correlation ratio and η^2 is the differentiation ratio [9, p. 137].

where δ is defined in Eq. (9.1) on p. 318 and μ_δ is the exact expected value of test statistic δ under the Fisher–Pitman null hypothesis given by

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

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

$$M = \left(g!\right)^{b-1} \ , \tag{9.13}$$

where g and b denote the number of treatments and blocks, respectively.

9.5.1 An Example Analysis

To illustrate the calculation of the five measures of effect size, suppose that a fast-food chain of restaurants decides to evaluate the service at four randomly-chosen restaurants. The customer-service director for the chain hires six evaluators with varied experiences in food-service evaluations to act as raters. In this example, the $g = 4$ restaurants are the treatments and the $b = 6$ raters are the blocks. The six raters evaluate the service at each of the four restaurants in random order. A rating scale from 0 (low) to 100 (high) is used. Table 9.6 summarizes the evaluation data.

Under the Neyman–Pearson population model with treatment means $\bar{x}_{1.} = 77.5000$, $\bar{x}_{2.} = 66.6667$, $\bar{x}_{3.} = 91.0000$, and $\bar{x}_{4.} = 79.3333$, block means $\bar{x}_{.1} = 71.7500$, $\bar{x}_{.2} = 79.0000$, $\bar{x}_{.3} = 78.2500$, $\bar{x}_{.4} = 78.7500$, $\bar{x}_{.5} = 81.5000$, and $\bar{x}_{.6} = 82.500$, grand mean $\bar{x}_{..} = 78.6250$, the sum-of-squares total is

$$SS_{\text{Total}} = \sum_{i=1}^{g} \sum_{j=1}^{b} \left(x_{ij} - \bar{x}_{..}\right)^2 = 2295.6250 \ ,$$

Table 9.6 Example restaurant data with $g = 4$ treatments and $b = 6$ blocks

	Restaurant			
Rater	A	B	C	D
1	70	61	82	74
2	77	75	88	76
3	76	67	90	80
4	80	63	96	76
5	84	66	92	84
6	78	68	98	86

the sum-of-squares treatments is

$$SS_{\text{Treatments}} = b \sum_{i=1}^{g} \left(\bar{x}_{i.} - \bar{x}_{..}\right)^2 = 1787.4583 \,,$$

the mean-square treatments is

$$MS_{\text{Treatments}} = \frac{SS_{\text{Treatments}}}{g-1} = \frac{1787.4583}{4-1} = 595.8194 \,,$$

the sum-of-squares blocks is

$$SS_{\text{Blocks}} = g \sum_{j=1}^{b} \left(\bar{x}_{.j} - \bar{x}_{..}\right)^2 = 283.3750 \,,$$

the sum-of-squares error is

$$\begin{aligned} SS_{\text{Error}} &= SS_{\text{Total}} - SS_{\text{Blocks}} - SS_{\text{Treatments}} \\ &= 2295.6250 - 283.3750 - 1787.4583 = 224.7917 \,, \end{aligned}$$

the mean-square error is

$$MS_{\text{Error}} = \frac{SS_{\text{Error}}}{(b-1)(g-1)} = \frac{224.7917}{(6-1)(4-1)} = 14.9861 \,,$$

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

$$F = \frac{MS_{\text{Treatments}}}{MS_{\text{Error}}} = \frac{595.8194}{14.9861} = 39.7581 \,.$$

The essential factors, sums of squares (SS), degrees of freedom (df), mean squares (MS), and variance-ratio test statistic (F) are summarized in Table 9.7.

Table 9.7 Source table for the data listed in Table 9.6

Factor	SS	df	MS	F
Blocks	283.3750			
Treatments	1787.4583	3	595.8194	39.7581
Error	224.7917	15	14.9861	
Total	2295.6250			

Given the summary data in Table 9.7, Hays' $\hat{\omega}^2$ measure of effect size is

$$\hat{\omega}^2 = \frac{(g-1)(MS_{\text{Treatments}} - MS_{\text{Error}})}{SS_{\text{Total}} + MS_{\text{Within Blocks}}} = \frac{(4-1)(595.8194 - 14.9861)}{2295.6250 - 111.7917} = 0.7238 ,$$

where the mean-square within blocks is

$$MS_{\text{Within Blocks}} = \frac{SS_{\text{Within Blocks}}}{b(g-1)} = \frac{2012.2500}{6(4-1)} = 111.7917 ,$$

and the sum-of-squares within blocks is

$$SS_{\text{Within Blocks}} = SS_{\text{Total}} - SS_{\text{Blocks}} = 2295.6250 - 283.3750 = 2012.2500 .$$

Following Eq. (9.9) on p. 327, Pearson's η^2 measure of effect size is

$$\eta^2 = \frac{SS_{\text{Treatments}}}{SS_{\text{Total}}} = \frac{1787.4583}{2295.6250} = 0.7786 .$$

Following Eq. (9.10) on p. 327, Cohen's partial η^2 measure of effect size is

$$\eta^2_{\text{Partial}} = \frac{SS_{\text{Treatments}}}{SS_{\text{Total}} - SS_{\text{Error}}} = \frac{1787.4583}{2295.6250 - 224.7917} = 0.8632 .$$

Following Eq. (9.11) on p. 327, Cohen's f^2 measure of effect size is

$$f^2 = \frac{SS_{\text{Treatments}}}{SS_{\text{Total}} - SS_{\text{Treatments}}} = \frac{1787.4583}{2295.6250 - 1787.4583} = 3.5175 .$$

Cohen's f^2 measure of effect size can also be defined in terms of Pearson's η^2 measure of effect size and calculated as

$$f^2 = \frac{\eta^2}{1 - \eta^2} = \frac{0.7786}{1 - 0.7786} = 3.5175 .$$

Following Eq. (9.13) on p. 328 with $\delta = 50.8167$,

$$M = \big(g!\big)^{b-1} = \big(4!\big)^{6-1} = 7962{,}624 ,$$

and following Eq. (9.6) on p. 320 the exact expected value of test statistic δ under the Fisher–Pitman null hypothesis is

$$\mu_\delta = \frac{1}{M}\sum_{i=1}^{M}\delta_i = \frac{1{,}560{,}873{,}370}{7{,}962{,}624} = 196.0250\,.$$

Then following Eq. (9.5) on p. 320, Mielke and Berry's chance-corrected measure of effect size is

$$\Re = 1 - \frac{\delta}{\mu_\delta} = 1 - \frac{50.8167}{196.0250} = +0.7408\,,$$

indicating approximately 78% within-blocks agreement above what is expected by chance.

A number of criticisms have been directed at the four conventional measures of effect size: Hays' $\hat{\omega}^2$, Pearson's η^2, Cohen's partial η^2, and Cohen's f^2. As can be seen in Eq. (9.8) on p. 327, in the unusual case when $MS_{\text{Treatments}}$ is smaller than MS_{Error}, yielding $F < 1$, Hays' $\hat{\omega}^2$ will be negative and it is difficult to interpret a squared measure of effect size that is negative. Moreover, unless a measure of effect size norms properly between the limits of 0 and 1, intermediate values are difficult to interpret.

Because Pearson's η^2 is simply the ratio of $SS_{\text{Treatments}}$ to SS_{Total}, η^2 norms properly between 0 and 1, providing an interpretation of the total variability in the dependent variable that is accounted for by variation in the independent variable. Moreover, when there is one degree of freedom in the numerator ($g = 2$ treatments), η^2 is equal to the product-moment coefficient of determination, r^2, and when there is more than one degree of freedom in the numerator ($g \geq 3$ treatments), η^2 is equal to the squared multiple product-moment correlation coefficient, R^2. Most researchers are familiar with Pearson's r^2 and R^2 correlation coefficients, making η^2 a useful index to understand the magnitude of effect sizes. Consequently, Pearson's η^2 is the most widely reported measure of effect size for randomized-blocks designs. On the other hand, η^2 is a biased estimator of effect size, systematically overestimating the size of treatment effects. Finally, as Sechrest and Yeaton concluded:

> As a general proposition it can be stated that *all measures of variance accounted for are specific to the characteristics of the experiment from which the estimates were obtained,* and therefore the ultimate interpretation of proportion of variance accounted for is a dubious prospect at best [10, p. 592].[3]

Cohen's partial η^2 is especially troublesome as reported by Kennedy [5], Levine and Hullett [6], Pedhazur [8, pp. 507–510], and Richardson [9]. In a classical one-way, completely-randomized analysis of variance design, η^2 and η^2_{Partial} yield identical results. However, η^2 and η^2_{Partial} yield different results in randomized-

[3] Emphasis in the original.

blocks analysis of variance designs, with η^2_{Partial} values being equal to or greater than η^2 values. Thus if η^2 systematically overestimates effect size, η^2_{Partial} overestimates effect size even more so. As Levine and Hullett concluded in reference to η^2:

> [B]ecause eta squared is always equal to partial eta squared or smaller, it may be seen as a more conservative estimate than partial eta squared and this may be appealing to many readers, reviewers, and editors [6, p. 620].

Since Cohen's η^2_{Partial} is not a percentage of the total sum-of-squares, it therefore is not additive like η^2. Moreover, η^2 has the advantage of being equivalent to the familiar r^2 and R^2 Pearson product-moment correlation coefficients.

Pedhazur pointed to another limitation of both η^2 and η^2_{Partial} as measures of effect size. While both η^2 and η^2_{Partial} have a logical upper bound of 1, the only situation in which η^2 and η^2_{Partial} can achieve an upper limit of 1 is when all values in each treatment are of one score, but differ among treatments. Pedhazur demonstrated that if the dependent variable is normally distributed, both η^2 and η^2_{Partial} have an upper limit of approximately 0.64 [8, p. 507]. Finally, Levine and Hullett concluded that "[O]ur examination of the literature revealed little reason for the reporting of partial eta squared" [6, p. 620].

Cohen's f^2 measure of effect size is seldom found in the literature as it is simply a function of Pearson's η^2. Cohen's f^2 is difficult to interpret as it varies between zero and infinity; for example, anytime Pearson's $\eta^2 > 0.50$, f^2 will exceed unity. Cohen suggested that small, medium, and large effects are reflected in values of f^2 equal to 0.01, 0.0625, and 0.16, respectively. In general, researchers desire more precision than simply small, medium, and large effect sizes.

On a more positive note, $\Re$ is a measure of effect size that possesses a clear and useful chance-corrected interpretation. Positive values of $\Re$ indicate agreement greater than expected by chance, negative values of $\Re$ indicate agreement less than expected by chance, and a value of zero indicates chance agreement. Moreover, $\Re$ is a universal measure of effect size and can be used in a wide variety of statistical applications, including one-sample t tests, matched-pairs t tests, simple and multiple regression, all manner of analysis of variance designs, and numerous contingency table analyses.

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

For a third example of tests of differences among $g \geq 3$ matched samples, consider the example data set given in Table 9.8 with $g = 3$ treatments, $b = 8$ blocks, and $N = bg = 24$ total observations. Under the Neyman–Pearson population model with treatment-group means $\bar{x}_{1.} = 229.25$, $\bar{x}_{2.} = 236.25$, and $\bar{x}_{3.} = 247.00$, block means $\bar{x}_{.1} = 241.00$, $\bar{x}_{.2} = 290.00$, $\bar{x}_{.3} = 118.6667$, $\bar{x}_{.4} = 246.3333$, $\bar{x}_{.5} = 122.6667$, $\bar{x}_{.6} = 336.00$, $\bar{x}_{.7} = 176.3333$, and $\bar{x}_{.8} = 369.00$, grand mean

Table 9.8 Example data for comparing analyses with $v = 2$ and $v = 1$ given $g = 3$ treatments and $b = 8$ blocks

	Treatment		
Block	1	2	3
1	221	247	255
2	283	302	285
3	103	130	123
4	254	223	262
5	115	113	140
6	322	344	342
7	161	181	187
8	375	350	382

$\bar{x}_{..} = 237.50$, the sum-of-squares total is

$$SS_{\text{Total}} = \sum_{i=1}^{g} \sum_{j=1}^{b} \left(x_{ij} - \bar{x}_{..}\right)^2 = 186{,}448.00 \,,$$

the sum-of-squares treatments is

$$SS_{\text{Treatments}} = b \sum_{i=1}^{g} \left(\bar{x}_{i.} - \bar{x}_{..}\right)^2 = 1279.00 \,,$$

the mean-square treatments is

$$MS_{\text{Treatments}} = \frac{SS_{\text{Treatments}}}{g - 1} = \frac{1279.00}{3 - 1} = 639.50 \,,$$

the sum-of-squares blocks is

$$SS_{\text{Blocks}} = g \sum_{j=1}^{b} \left(\bar{x}_{.j} - \bar{x}_{..}\right)^2 = 182{,}671.3333 \,,$$

the sum-of-squares error is

$$\begin{aligned} SS_{\text{Error}} &= SS_{\text{Total}} - SS_{\text{Blocks}} - SS_{\text{Treatments}} \\ &= 186{,}448.00 - 182{,}671.3333 - 1279.00 = 2497.6667 \,, \end{aligned}$$

the mean-square error is

$$MS_{\text{Error}} = \frac{SS_{\text{Error}}}{(b - 1)(g - 1)} = \frac{2497.6667}{(8 - 1)(3 - 1)} = 178.4048 \,,$$

Table 9.9 Source table for the data listed in Table 9.8

Factor	*SS*	*df*	*MS*	*F*
Blocks	182,671.3333			
Treatments	1279.0000	2	639.5000	3.5845
Error	2497.6667	14	178.4048	
Total	186,448.0000			

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

$$F = \frac{MS_{\text{Treatments}}}{MS_{\text{Error}}} = \frac{639.50}{178.4048} = 3.5845 \, .$$

The essential factors, sums of squares (*SS*), degrees of freedom (*df*), mean squares (*MS*), and variance-ratio test statistic (F) are summarized in Table 9.9.

Under the Neyman–Pearson null hypothesis, H_0: $\mu_1 = \mu_2 = \cdots = \mu_g$, Fisher's F-ratio test statistic is asymptotically distributed as Snedecor's F with $\nu_1 = g - 1$ and $\nu_2 = (b - 1)(g - 1)$ degrees of freedom. With $\nu_1 = g - 1 = 3 - 1 = 2$ and $\nu_2 = (b - 1)(g - 1) = (8 - 1)(3 - 1) = 14$ degrees of freedom, the asymptotic probability of $F = 3.5845$ is $P = 0.0553$, under the assumptions of normality and homogeneity.

9.6.1 An Exact Analysis with $v = 2$

For the example data listed in Table 9.8 with $g = 3$ treatments, $b = 8$ blocks, and $N = bg = 24$ observations, the observed value of the permutation test statistic with $v = 2$ is

$$\begin{aligned} \delta &= \frac{2[F SS_{\text{Blocks}} + (b - 1)SS_{\text{Total}}]}{g(b - 1)(F + b - 1)} \\ &= \frac{2[(3.5845)(182{,}671.3333) + (8 - 1)(186{,}448.00)]}{3(8 - 1)(3.5845 + 8 - 1)} = 17{,}635.1430 \, . \end{aligned}$$

Alternatively, in terms of a randomized-blocks analysis of variance model the observed permutation test statistic is

$$\begin{aligned} \delta &= \frac{2(SS_{\text{Total}} - SS_{\text{Treatments}})}{N - g} \\ &= \frac{2(186{,}448.00 - 1279.00)}{24 - 3} = 17{,}635.1430 \, . \end{aligned}$$

Because there are only

$$M = (g!)^{b-1} = (3!)^{8-1} = 279{,}936$$

possible, equally-likely arrangements in the reference set of all permutations of the observations listed in Table 9.8, an exact permutation analysis is 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 $N = 24$ observations listed in Table 9.8 that are equal to or less than the observed value of $\delta = 17{,}635.1430$. There are exactly 15,840 δ test statistic values that are equal to or less than the observed value of $\delta = 17{,}635.1430$. If all M arrangements of the $N = 24$ observations listed in Table 9.8 occur with equal chance under the Fisher–Pitman null hypothesis, the exact probability value computed on the $M = 279{,}936$ possible arrangements of the observed data with $b = 8$ blocks preserved for each arrangement is

$$P(\delta \leq \delta_o | H_0) = \frac{\text{number of } \delta \text{ values } \leq \delta_o}{M} = \frac{15{,}840}{279{,}936} = 0.0566\,,$$

where δ_o denotes the observed value of test statistic δ and M is the number of possible, equally-likely arrangements of the $N = 24$ observations listed in Table 9.8.

There are exactly 15,840 F values that are equal to or greater than the observed value of $F = 3.8582$. Thus, if all M arrangements of the $N = 24$ observations listed in Table 9.8 occur with equal chance under the Fisher–Pitman null hypothesis, the exact probability value of $F = 3.8582$ is

$$P(F \geq F_o | H_0) = \frac{\text{number of } F \text{ values } \geq F_o}{M} = \frac{15{,}840}{279{,}936} = 0.0566\,,$$

where F_o denotes the observed value of test statistic F.

Following Eq. (9.6) on p. 320, the exact expected value of the $M = 279{,}936$ δ test statistic values under the Fisher–Pitman null hypothesis is

$$\mu_\delta = \frac{1}{M}\sum_{i=1}^{M} \delta_i = \frac{4{,}958{,}224{,}209}{279{,}936} = 17{,}711.9921\,.$$

Following Eq. (9.5) on p. 320, the observed chance-corrected measure of effect size is

$$\Re = 1 - \frac{\delta}{\mu_\delta} = 1 - \frac{17{,}635.1430}{17{,}711.9921} = +0.4339 \times 10^{-2}\,,$$

indicating approximately chance within-block agreement.

9.6.2 Measures of Effect Size

Given the summary data in Table 9.9, Hays' $\hat{\omega}^2$ measure of effect size is

$$\hat{\omega}^2 = \frac{(g-1)(MS_{\text{Treatments}})}{SS_{\text{Total}} + MS_{\text{Within Blocks}}} = \frac{(3-1)(639.50-178.4048)}{186{,}448.00+236.0417} = 0.4940\times10^{-2},$$

where the mean-square within blocks is

$$MS_{\text{Within Blocks}} = \frac{SS_{\text{Within Blocks}}}{n(g-1)} = \frac{3776.6667}{8(3-1)} = 236.0417$$

and the sum-of-squares within blocks is

$$SS_{\text{Within Blocks}} = SS_{\text{Total}} - SS_{\text{Blocks}} = 186{,}448.00 - 182{,}671.3333 = 3776.6667.$$

Pearson's η^2 measure of effect size is

$$\eta^2 = \frac{SS_{\text{Treatments}}}{SS_{\text{Total}}} = \frac{1279.00}{186{,}448.00} = 0.6860\times10^{-2}.$$

Cohen's partial η^2 measure of effect size is

$$\eta^2_{\text{Partial}} = \frac{SS_{\text{Treatments}}}{SS_{\text{Total}} - SS_{\text{Error}}} = \frac{1279.00}{186{,}448.00 - 2497.6667} = 0.6953\times10^{-2}.$$

And Cohen's f^2 measure of effect size is

$$f^2 = \frac{SS_{\text{Treatments}}}{SS_{\text{Total}} - SS_{\text{Treatments}}} = \frac{1279.00}{186{,}448.00 - 1279.00} = 0.6813\times10^{-2}.$$

For comparison, Mielke and Berry's $\Re$ chance-corrected measure of effect size is

$$\Re = 1 - \frac{\delta}{\mu_\delta} = 1 - \frac{17{,}635.1430}{17{,}711.9921} = +0.4339\times10^{-2}.$$

In this case, the five measures of effect size yield about the same magnitude of experimental effect.

9.6.3 An Exact Analysis with $v = 1$

Following Eq. (9.1) on p. 318, for the example data listed in Table 9.8 on p. 333 with $g = 3$ treatments, $b = 8$ blocks, and $N = bg = 24$ observations, the observed value of the permutation test statistic with $v = 1$ is $\delta = 114.0238$. 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 $N = 12$ observations listed in Table 9.8 that are equal to or less than the observed value of $\delta = 114.0238$. There are exactly 172,986 δ test statistic values that are equal to or less than the observed value of $\delta = 114.0238$. If all M arrangements of the $N = 24$ observations listed in Table 9.8 occur with equal chance under the Fisher–Pitman null hypothesis, the exact probability value computed on the $M = 279{,}936$ possible arrangements of the observed data with $b = 8$ blocks preserved for each arrangement is

$$P(\delta \le \delta_o | H_0) = \frac{\text{number of } \delta \text{ values } \le \delta_o}{M} = \frac{163{,}296}{279{,}936} = 0.5833 ,$$

where δ_o denotes the observed value of test statistic δ and M is the number of possible, equally-likely arrangements of the $N = 24$ observations listed in Table 9.8. No comparison is made with Fisher's F-ratio test statistic as F is undefined for ordinary Euclidean scaling.

Following Eq. (9.6) on p. 320, the exact expected value of the $M = 279{,}936$ δ test statistic values under the Fisher–Pitman null hypothesis is

$$\mu_\delta = \frac{1}{M} \sum_{i=1}^{M} \delta_i = \frac{31{,}883{,}815}{279{,}936} = 113.8968 ,$$

and following Eq. (9.5) on p. 320, the observed chance-corrected measure of effect size is

$$\Re = 1 - \frac{\delta}{\mu_\delta} = 1 - \frac{114.0238}{113.8968} = -0.1114 \times 10^{-2} ,$$

indicating slightly less than chance within-block agreement. No comparisons are made with Hays' $\hat{\omega}^2$, Pearson's η^2, Cohen's partial η^2, or Cohen's f^2 measures of effect size as $\hat{\omega}^2$, η^2, η^2_{Partial}, and f^2 are undefined for ordinary Euclidean scaling.

9.6.4 The Effects of Extreme Values

To illustrate the robustness of ordinary Euclidean scaling with $v = 1$, consider the example data listed in Table 9.8 on p. 333 with changes made to the observations in Block 8. Suppose that an additional 20 points have been added to each of the

Table 9.10 Comparisons of exact permutation probability values with $v = 2$ and $v = 1$ for extreme block values

Change	Block 8	Probability	
		$v = 2$	$v = 1$
+0	375, 350, 382	0.057474	0.583333
+20	395, 370, 402	0.057474	0.583333
+40	415, 390, 422	0.057474	0.583333
+60	435, 410, 442	0.057474	0.583333
+80	455, 430, 462	0.057474	0.583333
+100	475, 450, 482	0.057474	0.583333
+120	495, 470, 502	0.057474	0.583333
+140	515, 490, 522	0.057474	0.583333
+160	535, 510, 542	0.057474	0.583333
+180	555, 530, 562	0.057474	0.583333
+200	575, 550, 582	0.057474	0.583333

$g = 3$ treatment values in Block 8. Block 8 contains the three largest values in each of the $g = 3$ treatments, making it the most extreme of all $b = 8$ blocks. The addition of 20 points increases the three values in Block 8 from {375, 350, 382} to {395, 370, 402}. A reanalysis of the data with the additional 20 points reveals that the probability values for $v = 2$ and $v = 1$ are unaffected by the extra 20 points. In fact, adding an additional 20 points (40 points total) does not alter the probability values. Table 9.10 illustrates the successive addition of 20 points, increasing up to an additional 200 points, demonstrating that the two permutation probability values remain constant. Thus tests under the Fisher–Pitman permutation model with both squared Euclidean scaling with $v = 2$ and ordinary Euclidean scaling with $v = 1$ are shown to be robust to an extreme block of data.

The same pattern holds with Fisher's F-ratio test statistic and asymptotic probability values. Table 9.11 lists the same block data as Table 9.10 with increments of 20 points added to the most extreme block, along with the associated F-ratio test statistic values and asymptotic probability values. The addition of extreme values

Table 9.11 Comparisons of Fisher's F-ratio test statistics and associated asymptotic probability values for extreme block values

Change	Block 8	F-ratio	Probability
+0	375, 350, 382	3.584545	0.055334
+20	395, 370, 402	3.584545	0.055334
+40	415, 390, 422	3.584545	0.055334
+60	435, 410, 442	3.584545	0.055334
+80	455, 430, 462	3.584545	0.055334
+100	475, 450, 482	3.584545	0.055334
+120	495, 470, 502	3.584545	0.055334
+140	515, 490, 522	3.584545	0.055334
+160	535, 510, 542	3.584545	0.055334
+180	555, 530, 562	3.584545	0.055334
+200	575, 550, 582	3.584545	0.055334

Table 9.12 Comparisons of exact permutation probability values with $v = 2$ and $v = 1$ for a single extreme value

Change	Block 8	Probability	
		$v = 2$	$v = 1$
+0	375, 350, 382	0.057474	0.583333
+20	375, 350, 402	0.040431	0.583333
+40	375, 350, 422	0.036912	0.583333
+60	375, 350, 442	0.032968	0.583333
+80	375, 350, 462	0.029635	0.583333
+100	375, 350, 482	0.027449	0.583333
+120	375, 350, 502	0.026299	0.583333
+140	375, 350, 522	0.025524	0.583333
+160	375, 350, 542	0.024945	0.583333
+180	375, 350, 562	0.024291	0.583333
+200	375, 350, 582	0.023823	0.583333

to a block does not change either the value of Fisher's F-ratio test statistic or the asymptotic probability value.

Now consider a different scenario. Suppose that an additional 20 points is added to only one treatment value in Block 8 in Table 9.8 on p. 333. The third value in Block 8 (382) is the largest of the $N = 24$ values. An additional 20 points increases value 382 to 402. In this case, the probability value based on ordinary Euclidean scaling with $v = 1$ is unchanged, remaining at $P = 0.583333$. However, the probability value based on squared Euclidean scaling with $v = 2$ decreases to $P = 0.040431$ from $P = 0.057474$. Table 9.12 illustrates the successive addition of 20 points, increasing up to an additional 200 points, demonstrating that ordinary Euclidean scaling with $v = 1$ under the Fisher–Pitman permutation model is robust to individual extreme values in randomized-blocks designs, while squared Euclidean scaling with $v = 2$ is not robust under the same model. The final probability value based on $v = 2$ of $P = 0.023823$ is less than half of the original probability value of $P = 0.057474$. The difference between the two exact probability values is

$$\Delta_P = 0.057474 - 0.023823 = 0.033651 .$$

For comparison, consider the block data listed in Table 9.13. The data listed in Table 9.13 are the same data listed in Table 9.12, but Table 9.13 also contains the F-ratio test statistic values and associated asymptotic probability values. As is clear from the results given in Table 9.13, Fisher's F-ratio test statistic values are strongly affected by the inclusion of a single extreme value in one block, as are the associated asymptotic probability values. The difference between the two F-ratio test statistics is

$$\Delta_F = 3.584545 - 2.162474 = 1.422071$$

and the difference between the two asymptotic probability values is

$$\Delta_P = 0.151914 - 0.055334 = 0.096580 .$$

Table 9.13 Comparisons of Fisher's F-ratio test statistic values and associated asymptotic probability values for a single extreme value

Change	Block 8	F-ratio	Probability
+0	375, 350, 382	3.584545	0.055334
+20	375, 350, 402	4.126205	0.039017
+40	375, 350, 422	4.097638	0.039726
+60	375, 350, 442	3.793197	0.048265
+80	375, 350, 462	3.434890	0.061132
+100	375, 350, 482	3.109992	0.076284
+120	375, 350, 502	2.838128	0.092321
+140	375, 350, 522	2.615944	0.108329
+160	375, 350, 542	2.434712	0.123762
+180	375, 350, 562	2.285893	0.138331
+200	375, 350, 582	2.162474	0.151914

9.7 Example 4: Exact and Monte Carlo Analyses

For a fourth example of tests for differences, consider the example data given in Table 9.14. It is generally understood that repeated experience with the Graduate Record Examination (GRE) leads to better scores, even without any intervening study. Suppose that eight subjects take the GRE verbal examination on successive Saturday mornings for three weeks. The data with $g = 3$ treatments, $b = 8$ blocks, and $N = 24$ scores are listed in Table 9.14.

Under the Neyman–Pearson population model with treatment means $\bar{x}_{1.} = 552.50$, $\bar{x}_{2.} = 564.3750$, and $\bar{x}_{3.} = 574.3750$, block means $\bar{x}_{.1} = 568.3333$, $\bar{x}_{.2} = 450.00$, $\bar{x}_{.3} = 616.6667$, $\bar{x}_{.4} = 663.3333$, $\bar{x}_{.5} = 436.6667$, $\bar{x}_{.6} = 696.6667$, $\bar{x}_{.7} = 505.00$, and $\bar{x}_{.8} = 573.3333$, grand mean $\bar{x}_{..} = 563.75$, the sum-of-squares total is

$$SS_{\text{Total}} = \sum_{i=1}^{g} \sum_{j=1}^{b} \left(x_{ij} - \bar{x}_{..}\right)^2 = 194{,}512.50 \,,$$

Table 9.14 Example GRE scores for exact and Monte Carlo analyses with $b = 8$ blocks and $g = 3$ treatments

	Treatment		
Block	1	2	3
1	550	575	580
2	440	440	470
3	610	630	610
4	650	670	670
5	400	460	450
6	700	680	710
7	490	510	515
8	580	550	590

the sum-of-squares treatments is

$$SS_{\text{Treatments}} = b \sum_{i=1}^{g} \left(\bar{x}_{i.} - \bar{x}_{..}\right)^2 = 1918.75 \,,$$

the mean-square treatments is

$$MS_{\text{Treatments}} = \frac{SS_{\text{Treatments}}}{g-1} = \frac{1918.75}{3-1} = 959.3750 \,,$$

the sum-of-squares blocks is

$$SS_{\text{Blocks}} = g \sum_{j=1}^{b} \left(\bar{x}_{.j} - \bar{x}_{..}\right)^2 = 189{,}112.50 \,,$$

the sum-of-squares error is

$$\begin{aligned} SS_{\text{Error}} &= SS_{\text{Total}} - SS_{\text{Blocks}} - SS_{\text{Treatments}} \\ &= 194{,}512.50 - 189{,}112.50 - 1918.75 = 3481.25 \,, \end{aligned}$$

the mean-square error is

$$MS_{\text{Error}} = \frac{SS_{\text{Error}}}{(b-1)(g-1)} = \frac{3481.25}{(8-1)(3-1)} = 248.6607 \,,$$

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

$$F = \frac{MS_{\text{Treatments}}}{MS_{\text{Error}}} = \frac{959.3750}{248.6607} = 3.8582 \,.$$

The essential factors, sums of squares (SS), degrees of freedom (df), mean squares (MS), and variance-ratio test statistic (F) are summarized in Table 9.15.

Under the Neyman–Pearson null hypothesis, H_0: $\mu_1 = \mu_2 = \cdots = \mu_g$, Fisher's F-ratio test statistic is asymptotically distributed as Snedecor's F with $\nu_1 = g - 1$ and $\nu_2 = (b-1)(g-1)$ degrees of freedom. With $\nu_1 = g - 1 = 3 - 1 = 2$ and $\nu_2 = (b-1)(g-1) = (8-1)(3-1) = 14$ degrees of freedom, the asymptotic probability value of $F = 3.8582$ is $P = 0.0463$, under the assumptions of normality and homogeneity.

Table 9.15 Source table for the GRE data listed in Table 9.13

Factor	SS	df	MS	F
Blocks	189,112.5000			
Treatments	1918.7500	2	959.3750	3.8582
Error	3481.2500	14	248.6607	
Total	194,512.5000			

9.7.1 An Exact Analysis with $v = 2$

For the first analysis of the data in Table 9.14 under the Fisher–Pitman permutation model let $v = 2$, employing squared Euclidean scaling for correspondence with Fisher's F-ratio test statistic. Because there are only

$$M = (g!)^{b-1} = (3!)^{8-1} = 279{,}936$$

possible, equally-likely arrangements in the reference set of all permutations of the $N = 24$ GRE scores listed in Table 9.14, an exact permutation analysis is feasible. Following Eq. (9.1) on p. 318, the observed value of the permutation test statistic is $\delta = 18{,}342.2620$. Based on the expressions given in Eqs. (9.3) and (9.4) on p. 319, the observed values of test statistics F and δ are

$$F = \frac{(b-1)[2SS_{\text{Total}} - g(b-1)\delta]}{g(b-1)\delta - 2SS_{\text{Blocks}}} = \frac{(8-1)[2(194{,}512.50) - 3(8-1)(18{,}342.2620)]}{3(8-1)(18{,}342.2620) - 2(189{,}112.50)} = 3.8582$$

and

$$\delta = \frac{2[FSS_{\text{Blocks}} + (b-1)SS_{\text{Total}}]}{g(b-1)(F+b-1)} = \frac{2[(3.8582)(189{,}112.50) + (8-1)(194{,}512.50)]}{3(8-1)(3.8582+8-1)} = 18{,}342.2620\,.$$

Alternatively, in terms of a randomized-blocks analysis of variance model the observed permutation test statistic is

$$\delta = \frac{2(SS_{\text{Total}} - SS_{\text{Treatments}})}{N-g} = \frac{2(194{,}512.50 - 1918.75)}{24-3} = 18{,}342.2620\,.$$

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 $N = 24$ observations listed in Table 9.14 that are equal to or less than the observed value of $\delta = 18{,}342.2620$. There are exactly 12,063 δ test statistic values that are equal to or less than the observed value of $\delta = 18{,}342.2620$. If all M arrangements of the $N = 24$ observations listed in Table 9.14 occur with equal chance under the Fisher–Pitman null hypothesis, the exact probability value of $\delta = 18{,}342.2620$ computed on the $M = 279{,}936$

possible arrangements of the observed data with $b = 8$ blocks preserved for each arrangement is

$$P(\delta \leq \delta_o | H_0) = \frac{\text{number of } \delta \text{ values } \leq \delta_o}{M} = \frac{12{,}063}{279{,}936} = 0.0431 ,$$

where δ_o denotes the observed value of test statistic δ and M is the number of possible, equally-likely arrangements of the GRE data listed in Table 9.14.

Alternatively, there are exactly 12,063 F-ratio test statistic values that are equal to or greater than the observed test statistic value of $F = 3.8582$. Thus, if all M arrangements of the $N = 24$ observations listed in Table 9.14 occur with equal chance under the Fisher–Pitman null hypothesis, the exact probability value of $F = 3.8582$ computed on the $M = 279{,}936$ arrangements of the observed data with $b = 4$ blocks preserved for each arrangement is

$$P(F \geq F_o | H_0) = \frac{\text{number of } F \text{ values } \geq F_o}{M} = \frac{12{,}063}{279{,}936} = 0.0431 ,$$

where F_o denotes the observed value of test statistic F.

Following Eq. (9.6) on p. 320, the exact expected value of the $M = 279{,}936$ δ test statistic values under the Fisher–Pitman null hypothesis is

$$\mu_\delta = \frac{1}{M} \sum_{i=1}^{M} \delta_i = \frac{5{,}167{,}818{,}515}{279{,}936} = 18{,}460.7143 .$$

Following Eq. (9.5) on p. 320, the observed chance-corrected measure of effect size is

$$\Re = 1 - \frac{\delta}{\mu_\delta} = 1 - \frac{18{,}342.2620}{18{,}460.7143} = +0.6416 \times 10^{-2} ,$$

indicating approximately chance within-block agreement.

9.7.2 Measures of Effect Size

For the GRE data listed in Table 9.14, Hays' $\hat{\omega}^2$ measure of effect size is

$$\hat{\omega}^2 = \frac{(g-1)(MS_{\text{Treatments}})}{SS_{\text{Total}} + MS_{\text{Within Blocks}}}$$
$$= \frac{(3-1)(959.3750 - 248.6607)}{194{,}512.00 + 337.50} = 0.7295 \times 10^{-2} ,$$

where the mean-square within blocks is

$$MS_{\text{Within Blocks}} = \frac{SS_{\text{Within Blocks}}}{n(g-1)} = \frac{5400.00}{8(3-1)} = 337.50$$

and the sum-of-squares within blocks is

$$SS_{\text{Within Blocks}} = SS_{\text{Total}} - SS_{\text{Blocks}} = 194{,}512.00 - 189{,}112.50 = 5400.00\,.$$

Pearson's η^2 measure of effect size is

$$\eta^2 = \frac{SS_{\text{Treatments}}}{SS_{\text{Total}}} = \frac{1918.75}{194{,}512.50} = 0.9864\times10^{-2}\,,$$

Cohen's partial η^2 measure of effect size is

$$\eta^2_{\text{Partial}} = \frac{SS_{\text{Treatments}}}{SS_{\text{Total}} - SS_{\text{Error}}} = \frac{1918.75}{194{,}512.00 - 3481.25} = 0.1004\times10^{-1}\,,$$

and Cohen's f^2 measure of effect size is

$$f^2 = \frac{SS_{\text{Treatments}}}{SS_{\text{Total}} - SS_{\text{Treatments}}} = \frac{1918.75}{194{,}512.50 - 1918.75} = 0.9963\times10^{-2}\,.$$

For comparison, Mielke and Berry's $\Re$ chance-corrected measure of effect size is

$$\Re = 1 - \frac{\delta}{\mu_\delta} = 1 - \frac{18{,}342.2620}{18{,}460.7143} = +0.6416\times10^{-2}\,.$$

Thus, the five measures of effect size yield about the same magnitude of experimental effect for this example analysis.

9.7.3 A Monte Carlo Analysis with $v = 2$

Although there are only $M = 279{,}936$ possible arrangements of the data listed in Table 9.14, making an exact permutation analysis feasible, many computer programs for permutation methods do not provide an option for an exact analysis. Moreover, over-sampling of the M possible arrangements is quite common in the permutation literature because of its efficiency in certain applications; for example, permutation analyses of contingency tables. In this section, over-sampling is demonstrated where $L = 1{,}000{,}000$ random arrangements is greater than the $M = 279{,}936$ possible arrangements.

For the example data listed in Table 9.14 on p. 340 with $v = 2$, the observed value of the permutation test statistic with $v = 2$ is $\delta = 18{,}342.2620$. 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 $N = 24$ observations listed in Table 9.14 that are equal to or less than the observed value of $\delta = 18{,}342.2620$. There are exactly 44,421 δ test statistic values that are equal to or less than the observed value of $\delta = 18{,}342.2620$. If all M arrangements of the $N = 24$ observations listed in Table 9.14 occur with equal chance under the Fisher–Pitman null hypothesis, the Monte Carlo probability value computed on a sample of $L = 1{,}000{,}000$ random arrangements of the observed data with $b = 8$ blocks preserved for each arrangement is

$$P(\delta \le \delta_o | H_0) = \frac{\text{number of } \delta \text{ values } \le \delta_o}{L} = \frac{44{,}421}{1{,}000{,}000} = 0.0444 ,$$

where δ_o denotes the observed value of test statistic δ and L is the number of randomly-selected, equally-likely arrangements of the GRE data listed in Table 9.14.

Alternatively, there are 44,421 F-ratio test statistic values that are equal to or greater than the observed value of $F = 3.8582$. Thus, if all M arrangements of the $N = 24$ observations listed in Table 9.14 occur with equal chance under the Fisher–Pitman null hypothesis, the Monte Carlo probability value of $F = 3.8582$ is

$$P(F \ge F_o | H_0) = \frac{\text{number of } F \text{ values } \ge F_o}{L} = \frac{44{,}421}{1{,}000{,}000} = 0.0444 ,$$

where F_o denotes the observed value of test statistic F.

The Monte Carlo probability value of $P = 0.0444$ based on $L = 1{,}000{,}000$ randomly-selected arrangements of the observed data compares favorably with the exact probability value of $P = 0.0431$ based on all $M = 279{,}936$ possible arrangements of the observed data.

Following Eq. (9.6) on p. 320, the exact expected value of the $M = 279{,}936$ δ test statistic values under the Fisher–Pitman null hypothesis is

$$\mu_\delta = \frac{1}{M} \sum_{i=1}^{M} \delta_i = \frac{5{,}167{,}818{,}515}{279{,}936} = 18{,}460.7143 ,$$

and following Eq. (9.5) on p. 320, the observed chance-corrected measure of effect size is

$$\Re = 1 - \frac{\delta}{\mu_\delta} = 1 - \frac{18{,}342.2620}{18{,}460.7143} = +0.6416 \times 10^{-2} ,$$

indicating approximately chance within-block agreement.

9.7.4 An Exact Analysis with $v = 1$

Consider a second analysis of the example data listed in Table 9.14 on p. 340 under the Fisher–Pitman permutation model with $v = 1$, employing ordinary Euclidean scaling between observations. For the data listed in Table 9.14 with $g = 3$ treatments, $b = 8$ blocks, and $N = bg = (8)(3) = 24$ observations, the observed permutation test statistic with $v = 1$ is $\delta = 116.3095$.

Because there are still only

$$M = (g!)^{b-1} = (3!)^{8-1} = 279{,}936$$

possible, equally-likely arrangements in the reference set of all permutations of the $N = 24$ GRE scores listed in Table 9.14, an exact permutation analysis is 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 $N = 24$ observations listed in Table 9.14 that are equal to or less than the observed value of $\delta = 116.3095$. There are exactly 186,624 δ test statistic values that are equal to or less than the observed value of $\delta = 116.3095$. If all M arrangements of the $N = 24$ observations listed in Table 9.14 occur with equal chance under the Fisher–Pitman null hypothesis, the exact probability value of $\delta = 116.3095$ computed on the $M = 279{,}936$ possible arrangements of the observed data with $b = 8$ blocks preserved for each arrangement is

$$P(\delta \leq \delta_o | H_0) = \frac{\text{number of } \delta \text{ values } \leq \delta_o}{M} = \frac{186{,}624}{279{,}936} = 0.6667\,,$$

where δ_o denotes the observed value of test statistic δ and M is the number of possible, equally-likely arrangements of the GRE data listed in Table 9.14. No comparison is made with Fisher's F-ratio test statistic as Fisher's F-ratio is undefined for ordinary Euclidean scaling.

Following Eq. (9.6) on p. 320, the exact expected value of the $M = 279{,}936$ δ test statistic values under the Fisher–Pitman null hypothesis is

$$\mu_\delta = \frac{1}{M}\sum_{i=1}^{M}\delta_i = \frac{32{,}514{,}790}{279{,}936} = 116.1508\,,$$

and following Eq. (9.5) on p. 320, the observed chance-corrected measure of effect size is

$$\Re = 1 - \frac{\delta}{\mu_\delta} = 1 - \frac{116.3095}{116.1508} = -0.1367\times10^{-2}\,,$$

indicating slightly less than chance within-block agreement. No comparisons are made with Hays' $\hat{\omega}^2$, Pearson's η^2, Cohen's partial η^2, or Cohen's f^2 measures of effect size as $\hat{\omega}^2$, η^2, η^2_{Partial}, and f^2 are undefined for ordinary Euclidean scaling.

For comparison, a Monte Carlo analysis based on $L = 1{,}000{,}000$ randomly-selected arrangements of the observed data listed in Table 9.14 with $v = 1$ yields $\delta = 116.3095$. 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 $N = 24$ observations listed in Table 9.14 that are equal to or less than the observed value of $\delta = 116.3095$. There are exactly 666,384 δ test statistic values that are equal to or less than the observed value of $\delta = 116.3095$. If all M arrangements of the $N = 24$ observations listed in Table 9.14 occur with equal chance under the Fisher–Pitman null hypothesis, the Monte Carlo probability value of $\delta = 116.3095$ computed on a sample of $L = 1{,}000{,}000$ randomly-selected arrangements of the observed data with $b = 8$ blocks preserved for each arrangement is

$$P(\delta \leq \delta_o | H_0) = \frac{\text{number of } \delta \text{ values } \leq \delta_o}{L} = \frac{666{,}384}{1{,}000{,}000} = 0.6664\,,$$

where δ_o denotes the observed value of test statistic δ and L is the number of randomly-selected, equally-likely arrangements of the GRE data listed in Table 9.14.

It is perhaps interesting that, for the example data listed in Table 9.14, the asymptotic probability value of $F = 3.8582$ with $\nu_1 = 2$ and $\nu_2 = 14$ degrees of freedom is $P = 0.0463$, the exact permutation probability value of $\delta = 18{,}342.2620$ with $v = 2$ is $P = 0.0431$, the Monte Carlo probability value of $\delta = 18{,}342.2620$ based on $L = 1{,}000{,}000$ random arrangements of the observed data is $P = 0.0444$, but the exact permutation probability value of $\delta = 116.3095$ with $v = 1$ is $P = 0.6667$. Thus the difference in exact probability values between analyses based on $v = 1$ and $v = 2$ is

$$\Delta_P = 0.6667 - 0.0431 = 0.6236\,,$$

which is a considerable discrepancy.

To be sure, the set of example data listed in Table 9.14 is rather innocuous—nothing unusual or extreme immediately presents itself. However, two values are somewhat extreme and it is extreme values that usually account for large differences in probability values based on squared Euclidean scaling with $v = 2$ and ordinary Euclidean scaling with $v = 1$. The two somewhat extreme values are $x_{6,1} = 700$ in Treatment 1 and $x_{6,3} = 710$ in Treatment 3. The value of 700 is 147.50 points above the average of Treatment 1 ($\bar{x}_{1\cdot} = 552.50$) and 1.42 standard deviations above the average value in Treatment 1. The value of 710 is 135.6250 points above the average of Treatment 3 ($\bar{x}_{3\cdot} = 574.3750$) and 1.48 standard deviations above the average value in Treatment 3.

The effects of these two values can be revealed by reducing the two values and re-analyzing the revised data. Consider reducing value $x_{6,1} = 700$ to $x_{6,1} = 600$, which with a standard score of $+0.46$ is closer to the mean of $\bar{x}_{.1} = 552.50$, and also reducing value $x_{6,3} = 710$ to $x_{6,3} = 600$, which with a standard score of $+0.28$ is closer to the mean of $\bar{x}_{.3} = 574.3750$. The result is to bring the probability values closer together, with an exact probability value based on squared Euclidean scaling with $v = 2$ of $P = 0.0799$, an exact probability value based on ordinary Euclidean scaling with $v = 1$ of $P = 0.2716$, and a difference between the two exact probability values of

$$\Delta_P = 0.2716 - 0.0799 = 0.1917$$

instead of a difference of

$$\Delta_P = 0.6667 - 0.0431 = 0.6236 .$$

The effects of the two extreme values can further be revealed by eliminating the two values. When the two values are eliminated—set equal to zero—and re-analyzed, the exact probability value based on squared Euclidean scaling with $v = 2$ is $P = 0.2651$ and the exact probability value based on ordinary Euclidean scaling with $v = 1$ is $P = 0.3914$ with a difference between the two exact probability values of only

$$\Delta_P = 0.3914 - 0.2651 = 0.1263 .$$

Table 9.16 lists the raw GRE scores from Table 9.14 on p. 340 along with associated standard scores, given in parentheses. To emphasize that the standard scores $+1.48$ and $+1.42$ are extreme relative to other scores listed in Table 9.16, a listing of the 13 positive standard scores in order is

$$\begin{aligned}\text{Standard score: } &+1.48,\ +1.42,\ +1.27,\ +1.27,\ +1.04,\ +0.94,\\ &+0.72,\ +0.55,\ +0.39,\ +0.27,\ +0.17,\ +0.16,\ +0.12 .\end{aligned}$$

Table 9.16 Example data from Table 9.14 with raw GRE scores and associated standard scores (in parentheses)

	Treatment		
Block	1	2	3
1	550 (−0.02)	575 (+0.12)	580 (+0.06)
2	440 (−1.09)	440 (−1.36)	470 (−1.14)
3	610 (+0.55)	630 (+0.72)	610 (+0.39)
4	650 (+0.94)	670 (+1.16)	670 (+1.04)
5	400 (−1.47)	460 (−1.14)	450 (−1.36)
6	700 (+1.42)	680 (+1.27)	710 (+1.48)
7	490 (−0.60)	510 (−0.52)	515 (−0.65)
8	580 (+0.27)	550 (−0.16)	590 (+0.17)

9.8 Example 5: Rank-Score Permutation Analyses

It is often necessary to analyze rank-score data when the required parametric assumptions of randomized-blocks designs cannot be met. However, with permutation methods it is never necessary to convert raw-score data to ranks [2]. The conventional approach to multi-sample rank-score data is Friedman's two-way analysis of variance for ranks [3].

9.8.1 The Friedman Analysis of Variance for Ranks

Let b denote the number of blocks and g denote the number of objects to be ranked. Then Friedman's test statistic is given by

$$\chi_r^2 = \frac{12}{bg(g+1)} \sum_{i=1}^{g} R_i^2 - 3b(g+1) \ ,$$

where R_i for $i = 1, \ldots, g$ is the sum of the rank scores for the ith object and there are no tied rank scores. A number of statistics are either identical, related, or equivalent to Friedman's χ_r^2 test statistic. Among these are Kendall and Babington Smith's coefficient of concordance, the average value of all pairwise Spearman's rank-order correlation coefficients, and the Wallis rank-order correlation ratio.

To illustrate Friedman's analysis of variance for ranks, consider the rank scores listed in Table 9.17; that is, rank scores r_{ij} for $i = 1, \ldots, g$ and $j = 1, \ldots, b$. For the rank-score data listed in Table 9.17, the sum of the squared rank scores is

$$\sum_{i=1}^{g} R_i^2 = 4^2 + 14^2 + 15^2 + 13^2 + 11^2 + 6^2 = 763 \ ,$$

Table 9.17 Example data for the Friedman analysis of variance for ranks with $b = 3$ blocks and $g = 6$ objects

	Block			
Object	1	2	3	R
1	1	1	2	4
2	6	5	3	14
3	3	6	6	15
4	4	4	5	13
5	5	2	4	11
6	2	3	1	6
Sum				63

and the observed value of Friedman's test statistic is

$$\chi_r^2 = \frac{12}{bg(g+1)} \sum_{i=1}^{g} R_i^2 - 3b(g+1)$$

$$= \frac{12}{(3)(6)(6+1)} 763 - (3)(3)(6+1) = 9.6667 .$$

Friedman's χ_r^2 test statistic is asymptotically distributed as Pearson's chi-squared under the Neyman–Pearson null hypothesis with $g - 1$ degrees of freedom. Under the Neyman–Pearson null hypothesis, the observed value of $\chi_r^2 = 9.6667$ with $g - 1 = 6 - 1 = 5$ degrees of freedom yields an asymptotic probability value of $P = 0.0853$.

9.8.2 An Exact Analysis with $v = 2$

For the first analysis of the rank-score data listed in Table 9.17 under the Fisher–Pitman permutation model let $v = 2$, employing squared Euclidean scaling between the pairs of rank scores for correspondence with Friedman's χ_r^2 test statistic, and let

$$x'_{ij} = (x_{1ij}, x_{2ij}, x_{3ij}, \ldots, x_{rij})$$

denote a transposed vector of r measurements associated with the ith treatment and jth block. Then the permutation test statistic is given by

$$\delta = \left[g \binom{b}{2} \right]^{-1} \sum_{i=1}^{g} \sum_{j=1}^{b-1} \sum_{k=j+1}^{b} \Delta(x_{ij}, x_{ik}) , \tag{9.14}$$

where $\Delta(x, y)$ is a symmetric distance-function value of two points $x' = (x_1, x_2, \ldots, x_r)$ and $y' = (y_1, y_2, \ldots, y_r)$ in an r-dimensional Euclidean space. In the context of a randomized-block design,

$$\Delta(x, y) = \sum_{i=1}^{r} |x_i - y_i|^v ,$$

where $v > 0$.

For the rank-score data listed in Table 9.17 there are only

$$M = (g!)^{b-1} = (6!)^{3-1} = 518{,}400$$

possible, equally-likely arrangements in the reference set of all permutations of the rank-score data listed in Table 9.17, making an exact permutation analysis feasible. For the rank scores listed in Table 9.17 let $v = 2$, employing squared Euclidean scaling between the pairs of rank scores for correspondence with Friedman's χ_r^2 test statistic, the observed value of the permutation test statistic with $v = 2$ is $\delta = 3.1111$.

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 9.17 that are equal to or less than the observed value of $\delta = 3.1111$. There are exactly 29,047 δ test statistic values that are equal to or less than $\delta = 3.1111$. If all M arrangements of the $N = 18$ rank scores listed in Table 9.17 occur with equal chance under the Fisher–Pitman null hypothesis, the exact probability of $\delta = 3.1111$ computed on the $M = 518{,}400$ possible arrangements of the observed rank scores with $b = 3$ blocks preserved for each arrangement is

$$P(\delta \leq \delta_o | H_0) = \frac{\text{number of } \delta \text{ values } \leq \delta_o}{M} = \frac{29{,}047}{518{,}400} = 0.0560\,,$$

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 9.17.

The functional relationships between test statistics χ_r^2 and δ with $v = 2$ are given by

$$\chi_r^2 = \frac{b(g^2 - 1) - 6(b - 1)\delta}{g + 1} \tag{9.15}$$

and

$$\delta = \frac{b(g^2 - 1) - (g + 1)\chi_r^2}{6(b - 1)}\,. \tag{9.16}$$

Following Eq. (9.15) for the $N = 18$ rank scores listed in Table 9.17, the observed value of test statistic χ_r^2 with respect to the observed value of test statistic δ is

$$\chi_r^2 = \frac{3(6^2 - 1) - 6(3 - 1)(3.1111)}{6 + 1} = 9.6667$$

and following Eq. (9.16), the observed value of test statistic δ with respect to the observed value of test statistic χ_r^2 is

$$\delta = \frac{3(6^2 - 1) - (6 + 1)(9.6667)}{6(3 - 1)} = 3.1111\,.$$

Following Eq. (9.6) on p. 320, the exact expected value of the $M = 518{,}400$ δ test statistic values under the Fisher–Pitman null hypothesis is

$$\mu_\delta = \frac{1}{M}\sum_{i=1}^{M}\delta_i = \frac{3{,}024{,}000}{518{,}400} = 5.8333\,.$$

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

$$\mu_\delta = \frac{2SS_{\text{Total}}}{N} = \frac{2(52.50)}{18} = 5.8333\,,$$

where

$$SS_{\text{Total}} = \sum_{i=1}^{g}\sum_{j=1}^{b} r_{ij}^2 - \left(\sum_{i=1}^{g}\sum_{j=1}^{b} r_{ij}\right)^2 \Big/ bg$$

$$= 273 - (63)^2/(3)(6) = 52.50\,.$$

Following Eq. (9.5) on p. 320, the observed chance-corrected measure of effect size is

$$\Re = 1 - \frac{\delta}{\mu_\delta} = 1 - \frac{3.1111}{5.8333} = +0.4667\,,$$

indicating approximately 47% within-block agreement above what is expected by chance. No comparisons are made with Hays' $\hat{\omega}^2$, Pearson's η^2, Cohen's partial η^2, or Cohen's f^2 measures of effect size as $\hat{\omega}^2$, η^2, η^2_{Partial}, and f^2 are undefined for rank-score data.

9.8.3 An Exact Analysis with $v = 1$

For a second analysis of the rank-score data listed in Table 9.17 under the Fisher–Pitman permutation model let $v = 1$, employing ordinary Euclidean scaling between the rank scores. For the rank scores listed in Table 9.17 there are still only

$$M = (g!)^{b-1} = (6!)^{3-1} = 518{,}400$$

possible, equally-likely arrangements in the reference set of all permutations of the rank-score data listed in Table 9.17, making an exact permutation analysis feasible. For the $N = 18$ rank scores listed in Table 9.17 the observed value of the permutation test statistic with $v = 1$ is $\delta = 1.4444$.

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 9.17 that are equal to or less than the observed value of $\delta = 1.4444$. There are exactly 55,528 δ test statistic values that are equal to or greater than $\delta = 1.4444$. If all M arrangements of the $N = 18$ rank scores listed in Table 9.17 occur with equal chance under the Fisher–Pitman null hypothesis, the exact probability of $\delta = 1.4444$ computed on the $M = 518{,}400$ possible arrangements of the observed rank scores with $b = 3$ blocks preserved for each arrangement is

$$P(\delta \leq \delta_o | H_0) = \frac{\text{number of } \delta \text{ values } \leq \delta_o}{M} = \frac{55{,}528}{518{,}400} = 0.1071 ,$$

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 9.17. No comparison is made with Friedman's χ_r^2 analysis of variance for ranks as χ_r^2 is undefined for ordinary Euclidean scaling.

Following Eq. (9.6) on p. 320, the exact expected value of the $M = 518{,}400$ δ test statistic values under the Fisher–Pitman null hypothesis is

$$\mu_\delta = \frac{1}{M} \sum_{i=1}^{M} \delta_i = \frac{1{,}008{,}000}{518{,}400} = 1.9444$$

and following Eq. (9.5) on p. 320, the observed chance-corrected measure of effect size is

$$\Re = 1 - \frac{\delta}{\mu_\delta} = 1 - \frac{1.4444}{1.9444} = +0.2571 ,$$

indicating approximately 26% within-block agreement above what is expected by chance. No comparisons are made with Hays' $\hat{\omega}^2$, Pearson's η^2, Cohen's partial η^2, or Cohen's f^2 measures of effect size as $\hat{\omega}^2$, η^2, η^2_{Partial}, and f^2 are undefined for rank-score data.

9.9 Example 6: Multivariate Permutation Analyses

It is oftentimes necessary to test for differences among $g \geq 3$ treatment groups where $r \geq 2$ measurements scores have been obtained from each of $b \geq 2$ blocks. To illustrate the analysis of randomized blocks with multivariate measurements, consider the data listed in Table 9.18 wherein each of two observers is asked to estimate distance and elevation in meters of 12 distant objects.

Table 9.18 Example data with $g = 12$ objects, $b = 2$ blocks, and $r = 2$ measurements

	Observer A		Observer B	
Object	Distance	Elevation	Distance	Elevation
1	120	10	125	10
2	80	15	85	20
3	100	5	95	10
4	150	20	140	15
5	75	10	60	5
6	50	5	60	10
7	50	20	50	25
8	20	20	25	15
9	90	15	90	15
10	95	25	90	20
11	100	25	90	20
12	70	5	70	5

9.9.1 A Monte Carlo Analysis with $v = 2$

For the example data listed in Table 9.18 with $g = 12$ treatments (objects), $b = 2$ blocks (observers), $r = 2$ measurements, and $N = bg = (2)(12) = 24$ multivariate observations, the observed value of the permutation test statistic with $v = 2$ is $\delta = 72.9167$. There are

$$M = (g!)^{b-1} = (12!)^{2-1} = 479{,}001{,}600$$

possible, equally-likely arrangements in the reference set of all permutations of the multivariate data listed in Table 9.18, making an exact permutation analysis impractical and a Monte Carlo analysis advisable. Under the Fisher–Pitman permutation model, the Monte Carlo probability value of an observed δ is the proportion of δ test statistic values computed on the randomly-selected, equally-likely arrangements of the $N = 24$ multivariate observations listed in Table 9.18 that are equal to or less than the observed value of $\delta = 72.9167$.

For the example data listed in Table 9.18 and $L = 1{,}000{,}000$ random arrangements of the observed data, there are exactly four δ test statistic values that are equal to or less than the observed value of $\delta = 72.9167$. If all M arrangements of the $N = 24$ observations listed in Table 9.18 occur with equal chance under the Fisher–Pitman null hypothesis, the Monte Carlo probability value of $\delta = 72.9167$ computed on $L = 1{,}000{,}000$ random arrangements of the observed data with $b = 2$ blocks preserved for each arrangement is

$$P(\delta \le \delta_o | H_0) = \frac{\text{number of } \delta \text{ values } \le \delta_o}{L} = \frac{4}{1{,}000{,}000} = 0.4000 \times 10^{-5} ,$$

where δ_o denotes the observed value of test statistic δ and L is the number of randomly-selected, equally-likely arrangements of the distance-elevation data listed in Table 9.18.

When the probability value is very small, as it is in this case, Monte Carlo permutation methods are not very precise with only $L = 1{,}000{,}000$ random arrangements of the observed data. A reanalysis of the multivariate data listed in Table 9.18 with $L = 100{,}000{,}000$ random arrangements yields a probability value of

$$P(\delta \leq \delta_o | H_0) = \frac{\text{number of } \delta \text{ values } \leq \delta_o}{L} = \frac{5}{100{,}000{,}000} = 0.5000 \times 10^{-7} .$$

Following Eq. (9.6) on p. 320, the exact expected value of test statistic δ under the Fisher–Pitman null hypothesis is

$$\mu_\delta = \frac{1}{M} \sum_{i=1}^{M} \delta_i = \frac{1{,}001{,}246{,}506{,}445}{479{,}001{,}600} = 2090.2780$$

and following Eq. (9.5) on p. 320, the observed chance-corrected measure of effect size is

$$\Re = 1 - \frac{\delta}{\mu_\delta} = 1 - \frac{72.9167}{2{,}090.2780} = +0.9651 ,$$

indicating approximately 97% within-block agreement above what is expected by chance. No comparisons are made with Hays' $\hat{\omega}^2$, Pearson's η^2, Cohen's partial η^2, or Cohen's f^2 measures of effect size as $\hat{\omega}^2$, η^2, η^2_{Partial}, and f^2 are undefined for multivariate data.

9.9.2 An Exact Analysis with $v = 2$

Although an exact permutation analysis with $M = 479{,}001{,}600$ possible arrangements of the observed data is not practical for the example data listed in Table 9.18, it is not impossible. For an exact permutation analysis with $v = 2$, the observed value of δ is $\delta = 72.9167$. There are exactly 20 δ test statistic values that are equal to or less than the observed value of $\delta = 72.9167$. If all M arrangements of the observed data occur with equal chance under the Fisher–Pitman null hypothesis, the exact probability value computed on the $M = 479{,}001{,}600$ possible arrangements of the observed data with $b = 2$ blocks preserved for each arrangement is

$$P(\delta \leq \delta_o | H_0) = \frac{\text{number of } \delta \text{ values } \leq \delta_o}{M} = \frac{20}{479{,}001{,}600} = 0.4175 \times 10^{-7} ,$$

where δ_o denotes the observed value of test statistic δ and M is the number of possible, equally-likely arrangements of the distance-elevation data listed in Table 9.18.

Following Eq. (9.6) on p. 320, the exact expected value of test statistic δ under the Fisher–Pitman null hypothesis is

$$\mu_\delta = \frac{1}{M}\sum_{i=1}^{M}\delta_i = \frac{0.1001\times10^{13}}{479{,}001{,}600} = 2090.2780$$

and following Eq. (9.5) on p. 320 the observed chance-corrected measure of effect size is

$$\Re = 1 - \frac{\delta}{\mu_\delta} = 1 - \frac{72.9167}{2090.2780} = +0.9651\ ,$$

indicating approximately 97% within-block agreement above what is expected by chance. No comparisons are made with Hays' $\hat{\omega}^2$, Pearson's η^2, Cohen's partial η^2, or Cohen's f^2 measures of effect size as $\hat{\omega}^2$, η^2, η^2_{Partial}, and f^2 are undefined for multivariate data.

9.9.3 A Monte Carlo Analysis with $v = 1$

For the data listed in Table 9.18 with $v = 1$, the observed value of δ is $\delta = 7.1305$. Since there are still

$$M = \big(g!\big)^{b-1} = \big(12!\big)^{2-1} = 479{,}001{,}600$$

possible, equally-likely arrangements in the reference set of all permutations of the multivariate data listed in Table 9.18, a Monte Carlo analysis is preferred. Under the Fisher–Pitman permutation model, the Monte Carlo probability value of an observed δ is the proportion of δ test statistic values computed on the randomly-selected, equally-likely arrangements of the $N = 24$ multivariate observations listed in Table 9.18 that are equal to or less than the observed value of $\delta = 7.1305$. For the data listed in Table 9.18 and $L = 1{,}000{,}000$ random arrangements of the data, there are exactly three δ test statistic values that are equal to or less than the observed value of $\delta = 7.1305$. If all M arrangements of the $N = 24$ multivariate observations listed in Table 9.18 occur with equal chance under the Fisher–Pitman null hypothesis, the Monte Carlo probability value of $\delta = 7.1305$ is

$$P(\delta \le \delta_o | H_0) = \frac{\text{number of } \delta \text{ values } \le \delta_o}{L} = \frac{3}{1{,}000{,}000} = 0.3000\times10^{-5}\ ,$$

where δ_o denotes the observed value of test statistic δ and L is the number of randomly-selected, equally-likely arrangements of the distance-elevation data listed in Table 9.18.

Following Eq. (9.6) on p. 320, the exact expected value of test statistic δ under the Fisher–Pitman null hypothesis is

$$\mu_\delta = \frac{1}{M}\sum_{i=1}^{M}\delta_i = \frac{17{,}916{,}053{,}734}{479{,}001{,}600} = 37.4029$$

and following Eq. (9.5) on p. 320, the observed chance-corrected measure of effect size is

$$\Re = 1 - \frac{\delta}{\mu_\delta} = 1 - \frac{7.1305}{37.4029} = +0.8094\,,$$

indicating approximately 81% within-block agreement above that is expected by chance. No comparisons are made with Hays' $\hat{\omega}^2$, Pearson's η^2, Cohen's partial η^2, or Cohen's f^2 measures of effect size as $\hat{\omega}^2$, η^2, η^2_{Partial}, and f^2 are undefined for multivariate data.

9.9.4 An Exact Analysis with $v = 1$

For an exact permutation analysis with $v = 1$, the observed value of δ is $\delta = 7.1305$. There are exactly four δ test statistic values that are equal to or less than the observed value of $\delta = 7.1305$. If all M arrangements of the $N = 24$ multivariate observations listed in Table 9.18 occur with equal chance under the Fisher–Pitman null hypothesis, the exact probability value of δ computed on the $M = 479{,}001{,}600$ possible arrangements of the observed data with $b = 2$ blocks reserved for each arrangement is

$$P(\delta \le \delta_o | H_0) = \frac{\text{number of } \delta \text{ values } \le \delta_o}{M}$$
$$= \frac{4}{479{,}001{,}600} = 0.8351\times10^{-8}\,,$$

where δ_o denotes the observed value of test statistic δ and M is the number of possible, equally-likely arrangements of the distance-elevation data listed in Table 9.18.

Following Eq. (9.6) on p. 320, the exact expected value of test statistic δ under the Fisher–Pitman null hypothesis is

$$\mu_\delta = \frac{1}{M}\sum_{i=1}^{M}\delta_i = \frac{17{,}916{,}053{,}734}{479{,}001{,}600} = 37.4029$$

and following Eq. (9.5) on p. 320 the observed chance-corrected measure of effect size is

$$\Re = 1 - \frac{\delta}{\mu_\delta} = 1 - \frac{7.1305}{37.4029} = +0.8094 ,$$

indicating approximately 81% within-block agreement above what is expected by chance. No comparisons are made with Hays' $\hat{\omega}^2$, Pearson's η^2, Cohen's partial η^2, or Cohen's f^2 measures of effect size as $\hat{\omega}^2$, η^2, η^2_{Partial}, and f^2 are undefined for multivariate data.

9.10 Summary

This chapter examined statistical methods for multiple dependent samples where the null hypothesis under the Neyman–Pearson population model posits no experimental 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 the conventional randomized-blocks analysis of variance and four measures of effect size were described and illustrated: Fisher's F test statistic, and Hays' $\hat{\omega}^2$, Pearson's η^2, Cohen's η^2_{Partial}, and Cohen's f^2 measures of effect size, respectively.

Under the Fisher–Pitman permutation model of statistical inference, test statistic δ and associated measure of effect size $\Re$ were described and illustrated for randomized-blocks designs. For tests of $g \geq 3$ dependent samples, test statistic δ was demonstrated to be applicable to both ordinary Euclidean scaling functions with $v = 1$ and squared Euclidean scaling functions with $v = 2$. Effect size measure, $\Re$, was shown to be applicable to either $v = 1$ or $v = 2$ without modification with a chance-corrected interpretation.

Six examples illustrated permutation-based test statistics δ and $\Re$ for randomized-blocks designs. In the first example, a small sample of $N = 8$ observations in $g = 2$ treatment groups and $b = 4$ blocks was utilized to describe and illustrate the calculation of test statistics δ and $\Re$ for randomized-blocks designs. The second example with $N = 24$ observations in $g = 4$ treatment groups and $b = 6$ blocks demonstrated the chance-corrected measure of effect size, $\Re$, for randomized-blocks designs and compared $\Re$ to the four conventional measures of effect size for $g \geq 3$ dependent samples: Hays' $\hat{\omega}^2$, Pearson's η^2, Cohen's partial η^2, and Cohen's f^2. The third example with $N = 24$ observations in $g = 3$ treatment groups and $b = 8$ blocks illustrated the effects of extreme values on analyses based on $v = 1$ for ordinary Euclidean scaling and $v = 2$ for squared Euclidean scaling. The fourth example with $N = 24$ observations in $g = 3$ treatment groups and $b = 8$ blocks compared exact and Monte Carlo permutation statistical methods for randomized-blocks designs, illustrating the accuracy and efficiency of Monte Carlo analyses. The fifth example with $N = 18$ observations in $g = 6$ treatment groups

and $b = 3$ blocks illustrated an application of permutation statistical methods to univariate rank-score data, comparing a permutation analysis of rank-score data with Friedman's g-sample analysis of variance for ranks. In the sixth example, both test statistic δ and effect-size measure $\Re$ were extended to multivariate data with $N = 48$ observations in $g = 12$ treatment groups, $b = 2$ blocks, and $r = 2$ measurements.

Chapter 10 continues the presentation of permutation statistical methods, examining permutation alternatives to simple linear correlation and regression. Research designs that utilize correlation and regression have a long history, are taught in every introductory class, and are among the most popular tests in the contemporary research literature.

References

1. Carroll, A.E.: A measured look at a study that alarmed some drinkers. N.Y. Times **167**, A12 (2018)
2. Feinstein, A.R.: Clinical biostatistics XXIII: the role of randomization in sampling, testing, allocation, and credulous idolatry (Part 2). Clin. Pharmacol. Ther. **14**, 898–915 (1973)
3. Friedman, M.: The use of ranks to avoid the assumption of normality implicit in the analysis of variance. J. Am. Stat. Assoc. **32**, 675–701 (1937)
4. Hotelling, H., Pabst, M.R.: Rank correlation and tests of significance involving no assumption of normality. Ann. Math. Stat. **7**, 29–43 (1936)
5. Kennedy, J.J.: The eta coefficient in complex ANOVA designs. Educ. Psych. Meas. **30**, 885–889 (1970)
6. Levine, T.R., Hullett, C.R.: Eta squared, partial eta squared, and misreporting of effect size in communication research. Hum. Commun. Res. **28**, 612–625 (2002)
7. Mielke, P.W., Berry, K.J.: Permutation Methods: A Distance Function Approach, 2nd edn. Springer, New York (2007)
8. Pedhazur, E.J.: Multiple Regression in Behavioral Research: Explanation and Prediction, 3rd edn. Harcourt, Fort Worth (1997)
9. Richardson, J.T.E.: Eta squared and partial eta squared as measures of effect size in educational research. Educ. Res. Rev. **6**, 135–147 (2011)
10. Sechrest, L., Yeaton, W.H.: Magnitude of experimental effects in social science research. Eval. Rev. **6**, 579–600 (1982)
11. Wood, A.M., Kaptage, S., Butterworth, A.S., Willeit, P., Warnakula, S., Bolton, T., et al.: Risk thresholds for alcohol consumption: combined analysis of individual-participant data for 599 912 current drinkers in 83 prospective studies. Lancet **391**, 1513–1523 (2018)