Completely Randomized Data

2

This second chapter of *Permutation Statistical Methods* introduces a generalized distance function that provides the foundation for a set of multi-response permutation procedures specifically designed for univariate and multivariate completely randomized data. Multi-Response Permutation Procedures (MRPP) were introduced by Mielke, Berry, and Johnson in 1976 and constitute a class of permutation methods for one or more response measurements on each object that were initially developed to distinguish possible differences among two or more groups of objects [300].[1] The multi-response permutation procedures presented here are based on a generalized Minkowski distance function and provide a synthesizing foundation for a variety of statistical tests and measures for completely randomized data that are further developed in Chaps. 3–7.

2.1 Minkowski Distance Function

Hermann Minkowski (1864–1909), German mathematician and creator of the geometry of numbers, utilized geometrical methods to solve problems in number theory, mathematical physics, and the theory of relativity. Minkowski was a close friend of David Hilbert while teaching at Königsberg University and taught Albert Einstein while employed at Eidgenössische Polytechnikum in Zürich (now, ETH Zürich). In 1891 Minkowski introduced a measure of metric distance between

---

[1] The 1976 paper by Mielke, Berry, and Johnson was the first published account of MRPP [300]. Previously, Mielke utilized MRPP in a study sponsored by the National Communicable Disease Center that involved comparisons of proportional contributions of five plague organism protein bands based on electrophoresis measurements obtained from samples of organisms associated with distinct geographical regions.

K.J. Berry et al., *Permutation Statistical Methods*,
DOI 10.1007/978-3-319-28770-6_2

two points in *Crelle's Journal* [310].[2] The Minkowski metric distance of order p between two points in an r-dimensional Euclidean space, $x' = (x_1, x_2, \ldots, x_r)$ and $y' = (y_1, y_2, \ldots, y_r) \in \mathbb{R}^r$, is given by

$$d(x,y) = \left(\sum_{i=1}^{r} \left| x_i - y_i \right|^p \right)^{1/p} ,$$

where $p \geq 1$.

The Minkowski distance function is typically used with $p = 1, 2$, or ∞. When $p = 1$, the distance is a first-order Minkowski metric, often called a city-block, Manhattan [231], rectilinear [54], or taxicab [222] metric, the latter named for the distance between two points that a car or taxicab would drive in a city laid out in square blocks. When $p = 2$, the distance is a second-order Minkowski metric and is the ordinary Euclidean distance between points, a generalization of the Pythagorean theorem to more than two coordinates. When $p = \infty$, the Minkowski metric is known as the Tchebycheff (Chebyshev), von Neumann, or, in the two-dimensional case, the chess-board Minkowski distance [167].

Conventional statistical tests and measures, such as t tests, F tests, and ordinary least-squares (OLS) regression and correlation, are based on squared Euclidean distances between response measurement scores, which are not metric. The Minkowski distance function, however, is limited to metric distances and, under its standard definition, cannot accommodate most conventional statistical tests. Therefore, consider a generalized Minkowski distance function given by

$$\Delta(x,y) = \left(\sum_{i=1}^{r} \left| x_i - y_i \right|^p \right)^{v/p} , \tag{2.1}$$

where $p \geq 1$ and $v > 0$ [297, p. 5]. When $r \geq 2$, $p = 2$, and $v = 1$, $\Delta(x,y)$ is rotationally invariant in an $r \geq 2$ dimensional space. When $v = p = 1$, $\Delta(x,y)$ is a city-block metric, which is not rotationally invariant. When $v = 1$ and $p = 2$, $\Delta(x,y)$ is an ordinary Euclidean distance metric. And when $v = p = 2$, $\Delta(x,y)$ is a squared Euclidean distance, which is not a metric distance function since the triangle inequality is not satisfied.[3]

---

[2] The *Journal für die Reine und Angewandte Mathematik* was founded by August Leopold Crelle in 1826. It continues today, although it is more popularly known as *Crelle's Journal*.

[3] A distance function is a metric if it satisfies three properties given by (1) $\Delta(x,y) \geq 0$ and $\Delta(x,x) = 0$, i.e., the distance is positive between two different points and is equal to zero from any point to itself; (2) the distance is symmetric: $\Delta(x,y) = \Delta(y,x)$, i.e., the distance between points x and y is the same in either direction; and (3) the triangle inequality is satisfied: $\Delta(x,y) \leq \Delta(x,z) + \Delta(z,y)$, i.e., the distance between any two points is the shortest distance along any path.

2.2 Multi-response Permutation Procedures

Multi-Response Permutation Procedures (MRPP) were originally designed to statistically determine possible differences among one or more response measurement scores among two or more groups of objects or subjects [300]. Let $\Omega = \{\omega_1, \ldots, \omega_N\}$ denote a finite sample of N objects that represents a target population, let $x_i' = (x_{1i}, \ldots, x_{ri})$ be a transposed vector of r commensurate response measurement scores for object ω_i, $i = 1, \ldots, N$, and let $S_1, \ldots, S_g$ designate an exhaustive partitioning of the N objects into g disjoint treatment groups.[4] The MRPP test statistic is a weighted mean given by

$$\delta = \sum_{i=1}^{g} C_i \xi_i \;, \tag{2.2}$$

where $C_i > 0$ is a positive weight for treatment group S_i, $i = 1, \ldots, g$, $\sum_{i=1}^{g} C_i = 1$,

$$\xi_i = \binom{n_i}{2}^{-1} \sum_{j<k} \Delta(j,k)\, \Psi_i(\omega_j)\, \Psi_i(\omega_k) \tag{2.3}$$

is the average distance-function value for all distinct pairs of objects in treatment group S_i, $i = 1, \ldots, g$, $n_i \geq 2$ is the number of a priori objects classified into treatment group S_i, $i = 1, \ldots, g$,

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

$\sum_{j<k}$ is the sum over all j and k such that $1 \leq j < k \leq N$, and $\Psi(\cdot)$ is an indicator function given by

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

The choice of the treatment-group weights, $C_1, \ldots, C_g$, and the generalized Minkowski distance function given in Eq. (2.1) on p. 30 specify the structure of

---

[4] Multi-response permutation procedures also provide for a group of unclassified response measurement scores such as might result from a survey with question choices that include "none of the above" or "not applicable." See, for example, a 1983 article on lead concentrations in inner-city soils by Mielke, Anderson, Berry, Mielke, Chaney, and Leech [302] and a discussion by Mielke and Berry in 2007 [297, pp. 35–40].

MRPP. The original choice of C_i given by Mielke, Berry, and Johnson in 1976 was

$$C_i = \frac{n_i(n_i - 1)}{\sum_{j=1}^{g} n_j(n_j - 1)}$$

for $i = 1, \ldots, g$ [300]. However, a variety of other treatment-group weights can be considered; for example,

$$C_i = \frac{n_i}{N}, \quad C_i = \frac{n_i - 1}{N - g}, \quad \text{or} \quad C_i = \frac{1}{g}$$

for $i = 1, \ldots, g$. The efficient choice of $C_i = n_i/N$, $i = 1, \ldots, g$, forces the population variance, σ_x^2, to be proportional to N^{-2} and eliminates all terms of order $1/N$ in the variance of δ [297, pp. 26, 30].

The null hypothesis (H_0) states that equal probabilities are assigned to each of the

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

possible, equally-likely allocations of the N objects to the g treatment groups, $S_1, \ldots, S_g$. Under H_0 the N multi-response measurements are exchangeable multivariate random variables.[5] The probability value associated with an observed value of δ, δ_o, is the probability under the null hypothesis (H_0) of observing a value of δ as extreme or more extreme than δ_o. Thus, an exact probability value for δ_o may be expressed as

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

When M is very large, an approximate probability value for δ may be obtained from a resampling procedure, where

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

---

[5] A sufficient condition for a permutation statistical test is the exchangeability of the random variables. Sequences that are independent and identically distributed (i.i.d.) are always exchangeable, but so is sampling without replacement from a finite population. However, while i.i.d. implies exchangeability, exchangeability does not imply i.i.d. [150, 168, 217].

and L denotes the number of randomly sampled test statistic values. Typically, L is set to a large number to ensure accuracy.

Number of Resamplings Necessary

Exact permutation tests are restricted to relatively small samples, given the large number of possible permutations. On the other hand, resampling permutation tests are not limited by the size of the samples. Resampling permutation tests also have been shown to provide good approximations to exact probability values as a function of the number of resamplings considered. An early concern regarding the systematic use of resampling permutation tests was the speed of the computers used for calculating the probability values. Given modern high-speed computers, the question of computational speed is moot when probability values are not too small. The remaining question is: how many resamplings are required for a specified accuracy?

The number of resamplings suggested in books and articles on permutation methods is varied and likely dated due to previous limitations of computer speed and memory. Some authors have proposed as few as 100 resamplings to as many as 5,000; for example, see discussions by Dwass in 1957 [100]; Hope in 1968 [180]; Edwards in 1985 [110]; Jockel in 1986 [193]; Keller-McNulty and Higgins in 1987 [199]; Bailer in 1989 [16]; Kim, Nelson, and Startz in 1991 [216]; Manly in 1991 [258, pp. 32–35]; McQueen in 1992 [274]; Rickerts and Berry in 1994 [347]; Kennedy in 1995 [212]; Maxim in 1999 [265, p. 356]; Lunneborg in 2000 [256, pp. 210–213]; Good in 2001 [149, p. 47]; Higgins in 2004 [176]; and Edgington and Onghena in 2007 [109, pp. 40–41]. On the other hand, examples provided by Howell as recently as 2007 utilized as many as 10,000 resamplings [184, pp. 642–646]. Resampling computing packages such as Resampling Stats [14] and StatXact [15] typically use 10,000 resamplings as the default value.

The accuracy of a resampling probability value depends on both the probability value (P) and the number of resamplings (L). Confidence limits on the probability value can be obtained from the binomial distribution when L is large. The $1-\alpha$ confidence limits of the binomial distribution are given by

$$\hat{P} \pm Z_{\alpha/2} \sqrt{\frac{P(1-P)}{L}} \,, \tag{2.4}$$

where P is the probability value in question and $\hat{P}$ denotes the estimated value of P. Define

$$x_i = \begin{cases} 1 & \text{if } \hat{P} \le \hat{P}_{\text{o}} \,, \\ 0 & \text{otherwise} \,, \end{cases}$$

for $i = 1, \ldots, L$, where $\hat{P}_{\text{o}}$ denotes the observed value of $\hat{P}$. Then $\hat{P}$, the expected value of $\hat{P}$, the variance of $\hat{P}$, and the skewness of $\hat{P}$ are given by

$$\hat{P} = \frac{1}{L} \sum_{i=1}^{L} x_i \ ,$$

$$\mathrm{E}[\hat{P}] = P \ ,$$

$$\sigma_{\hat{P}}^2 = \frac{P(1-P)}{L} \ ,$$

and

$$\gamma_{\hat{P}} = \frac{1-2P}{\sqrt{LP(1-P)}} \ ,$$

respectively [195, p. 916]. If L is small and P is close to either 0 or 1, the skewness term $\gamma_{\hat{P}}$ becomes large and Eq. (2.4) may not be appropriate. For example, if $L = 100$ and $P = 0.01$,

$$\gamma_{\hat{P}} = \frac{1-2P}{\sqrt{LP(1-P)}} = \frac{1-2(0.01)}{\sqrt{100(0.01)(1-0.01)}} = 0.9849 \ .$$

Table 2.1 lists a selected number of probability values ($P = 0.50$, 0.25, 0.10, 0.05, and 0.01), a variety of resamplings ($L = 100$, 1000, 10,000, 1,000,000, and 100,000,000), computed skewness values, errors on the 95 % confidence limits determined from Eq. (2.4), and the simulated lower and upper errors on the 95 % confidence limits based on L resamplings and determined from the smallest value for which the cumulative binomial distribution is equal to or less than 0.025 and equal to or greater than 0.975, respectively. In general, as can be seen from Table 2.1, two additional orders of magnitude are required to increase accuracy by just one decimal place.

To illustrate the number of resamplings required to yield a predetermined number of decimal places of accuracy, given a known probability value, consider the interval-level data listed in Fig. 2.1.

The data listed in Fig. 2.1 are adapted from Berry, Mielke, and Mielke [38] and represent soil lead (Pb) quantities from two school districts in metropolitan New Orleans. Elevated Pb levels have been linked to a number of physiological, neurological, and endocrine effects in children, including difficulties in learning, perception, social behavior, and fine motor skills. The $n_1 = 20$ soil lead samples collected in District 1 yielded a mean value of $\bar{x}_1 = 203.9350$ mg/kg and the $n_2 = 20$ soil lead samples collected in District 2 yielded a mean value of $\bar{x}_2 = 1{,}661.7800$ mg/kg. There are

$$M = \frac{(n_1+n_2)!}{n_1!\, n_2!} = \frac{(20+20)!}{20!\, 20!} = 137{,}846{,}528{,}820$$

Table 2.1 Five probability (P) values, four levels of resampling (L), skewness ($\gamma_{\hat{P}}$), and asymptotic and simulated errors on 95 % confidence limits; table adapted from Johnston, Berry, and Mielke [195, p. 917]

			Error on 95 % confidence limits		
P	L	$\gamma_{\hat{P}}$	Asymptotic	Lower	Upper
0.50	100	0.00	0.10	0.10	0.10
	10,000	0.00	0.010	0.010	0.010
	1,000,000	0.00	0.0010	0.0010	0.0010
	100,000,000	0.00	0.00010	0.00010	0.00010
0.25	100	0.11547	0.09	0.08	0.09
	10,000	0.01155	0.009	0.009	0.009
	1,000,000	0.00115	0.0009	0.0008	0.0008
	100,000,000	0.00012	0.00009	0.00008	0.00008
0.10	100	0.26667	0.06	0.05	0.06
	10,000	0.02667	0.006	0.006	0.006
	1,000,000	0.00267	0.0006	0.0006	0.0006
	100,000,000	0.00027	0.00006	0.00006	0.00006
0.05	100	0.41295	0.04	0.04	0.05
	10,000	0.04129	0.004	0.004	0.004
	1,000,000	0.00413	0.0004	0.0004	0.0004
	100,000,000	0.00041	0.00004	0.00004	0.00004
0.01	100	0.98494	0.02	0.01	0.02
	10,000	0.09849	0.002	0.002	0.002
	1,000,000	0.00985	0.0002	0.0002	0.0002
	100,000,000	0.00098	0.00002	0.00002	0.00002

possible permutations of the soil lead data listed in Fig. 2.1 to be considered. Under the null hypothesis of no difference between the two group means in the population, a Fisher–Pitman permutation F test [38] yields an exact two-sided probability value of

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

$$= \frac{2{,}056{,}423{,}782}{137{,}846{,}528{,}820} = 0.0149182123$$

for the soil lead data listed in Fig. 2.1. Figure 2.2 summarizes the results for eight different resamplings of the data listed in Fig. 2.1 and the associated two-sided resampling probability values with $\alpha = 0.05$. Each of the probability values was generated using a common seed and the same pseudorandom number generator [197]. The last row of Fig. 2.2 contains the exact probability value based on all $M = 137{,}846{,}528{,}820$ possible permutations of the soil lead data listed in Fig. 2.1.

Fig. 2.1 Ordered soil Pb data in mg/kg from two school attendance districts in metropolitan New Orleans

n	District 1	District 2
1	16.0	4.7
2	34.3	10.8
3	34.6	35.7
4	57.6	53.1
5	63.1	75.6
6	88.2	105.5
7	94.2	200.4
8	111.8	212.8
9	112.1	212.9
10	139.0	215.2
11	165.6	257.6
12	176.7	347.4
13	216.2	461.9
14	221.1	566.0
15	276.7	984.0
16	362.8	1,040.0
17	373.4	1,306.0
18	387.1	1,908.0
19	442.2	3,559.0
20	706.0	21,679.0

Fig. 2.2 Comparison of eight resampled probability values with the exact probability value given in the last row, based on the soil lead data listed in Fig. 2.1

Resampling (L)	Probability ($\hat{P}$)
10	0.06
1,000	0.020
10,000	0.0110
100,000	0.01556
1,000,000	0.014946
10,000,000	0.0149302
100,000,000	0.01488510
1,000,000,000	0.014917218
Exact P value	0.0149182123

Given the results of the resampling probability analyses listed in Fig. 2.2, $L =$ 1,000,000 is recommended whenever three decimal places of accuracy are required. There are four reasons for promoting $L =$ 1,000,000 resamplings: accuracy, practicality, error, and consistency. First, inspection of Fig. 2.2 indicates that with an exact probability value of $P = 0.0149182123$ and $\alpha = 0.05$, $L =$ 1,000,000 resamplings is the minimum number of resamplings necessary to ensure three decimal places of accuracy. Second, given the speed of modern computers and the efficiency of resampling algorithms, such as the Mersenne Twister, $L =$ 1,000,000 resamplings can be used on a routine basis. Third, there is the potential for additional type I error, the magnitude of which is of concern when the number of resamplings (L) is very small. Fourth, some researchers object to the use of resampling statistics because different pseudorandom number generators and different seeds can produce widely varying results. This is certainly true when L is very small. For example, in Fig. 2.2, $L = 100$ yields a probability value of $P = 0.06$. Varying the seed with

$L = 100$ and the same pseudorandom number generator produced observed probability values ranging from $P = 0.01$ to $P = 0.11$. However, with $L = 1{,}000{,}000$, varying the seed produced no differences in the third decimal place.

When the number of possible arrangements (M) is very large and the exact probability value (P) is exceedingly small, a resampling permutation procedure may produce no δ values equal to or less than δ_{o}, even with $L = 1{,}000{,}000$, yielding an approximate resampling probability value of $P = 0.00$. In such cases, moment-approximation permutation procedures based on fitting the first three exact moments of the discrete permutation distribution to a Pearson type III distribution provide approximate probability values, as detailed in Chap. 1, Sect. 1.2.2; see also references [284] and [300].

An Index of Agreement

It is oftentimes desirable to have an index of the amount of agreement among response measurement scores within g treatment groups. A useful measure for this purpose is a chance-corrected within-group coefficient of agreement given by

$$\Re = 1 - \frac{\delta}{\mu_\delta} \,, \tag{2.5}$$

where μ_δ is the arithmetic average of the M δ values calculated on all possible arrangements of the observed response measurement scores, given by

$$\mu_\delta = \frac{1}{M} \sum_{i=1}^{M} \delta_i \,. \tag{2.6}$$

$\Re$ is a chance-corrected measure of agreement since $\mathrm{E}[\Re|H_0] = 0$.[6] Because μ_δ is a constant under H_0, the permutation distributions of δ and $\Re$ are equivalent, viz.,

$$P\big(\delta \leq \delta_{\text{o}}|H_0\big) = P\left(\Re \geq \Re_{\text{o}}|H_0\right) \,,$$

where

$$\Re_{\text{o}} = 1 - \frac{\delta_{\text{o}}}{\mu_\delta}$$

and δ_{o} and $\Re_{\text{o}}$ denote the observed values of δ and $\Re$, respectively. Possible values of $\Re$ range from slightly negative values to a maximum of $\Re = +1$ for the extreme

---

[6] As will be shown in Chap. 3, $\Re$ may also be interpreted as a chance-corrected measure of effect size.

case when all response measurements on objects within each of the g classified treatment groups are identical, i.e., $\delta = 0$.

The generalized Minkowski distance function, $\Delta(x, y)$, as defined in Eq. (2.1) on p. 30, determines the analysis space of the MRPP test statistic, δ. The data space in question for almost all statistical analyses is an ordinary Euclidean distance space. If the distance function of the MRPP test statistic is based on $p = 2$ and $v = 1$, then the data and analysis spaces are congruent, so that the resulting statistical analyses represent the data in question. Unfortunately, commonly used statistical analyses based on the arithmetic mean, such as Student's two-sample t test and Fisher's one-way analysis of variance, are based on $p = v = 2$, yielding a non-metric squared-distance analysis space that is not congruent with the data space. The difference between the data and analysis spaces associated with the most popular statistical analyses is a reason that problems occur with what should be routine analyses. Examples illustrating this problem are given elsewhere; see, for example, references [41, pp. 404–410] and [297, pp. 50–53]. Any statistical analysis is questionable when the data and analysis spaces are not congruent.

2.2.1 Chance-Corrected Agreement Measures

Chance-corrected measures yield values that are interpreted as a proportion above that expected by chance alone. Chance-corrected agreement measures provide clear and meaningful interpretations of the amount of, or lack of, agreement present in the data. In general, chance-corrected measures of agreement, such as $\Re$, are equal to $+1$ when perfect agreement among the response measurement scores occurs, 0 when agreement is equal to that expected under independence, and negative when agreement among the response measurement scores is less than that expected by chance. For example, define a chance-corrected measure such that

$$A_i = 100 \left(\frac{O_i - E_i}{N - E_i} \right) ,$$

where O_i and E_i denote the Observed (earned) and Expected (chance) score from purely guessing, respectively, on a multiple-choice examination with N questions for the ith student in a class of m students [175, p. 912].

Thus, on a 50-question multiple-choice examination with five choices per question, chance would indicate that a student could answer $50 \times 0.20 = 10$ questions correctly simply by guessing. If a student answered only eight questions correctly, then a chance-corrected measure of agreement would yield a grade of

$$A = 100 \left(\frac{8 - 10}{50 - 10} \right) = 100 \left(\frac{-2}{40} \right) = -5 ,$$

since the score was less than expected by chance, i.e., only eight of 50 questions were answered correctly. The lowest grade would occur when a student answered all 50 questions incorrectly, yielding a score of

$$A = 100\left(\frac{0-10}{50-10}\right) = 100\left(\frac{-10}{40}\right) = -25\ .$$

Note that while a student with the highest possible score of 50 correct answers would score

$$A = 100\left(\frac{50-10}{50-10}\right) = 100\left(\frac{40}{40}\right) = 100\ ,$$

the lowest possible score is −25, not −100. Thus, the distributions of chance-corrected measures are usually asymmetric.

Since the mean value of $\Re$ under H_0 is 0, homogeneity of within-classified-group response measurements is associated with $\Re > 0$, and heterogeneity of within-classified-group response measurements is associated with $\Re \leq 0$ [28]. The distribution of $\Re$ is usually asymmetric and the upper and lower bounds depend on both the nature of the data and the structure of δ. The degree of homogeneity or heterogeneity depends on the discrete permutation distribution of $\Re$. If large values of $n_1, \ldots, n_g$ and N are involved, a very small value of $P(\delta \leq \delta_o | H_0)$ may be associated with a small positive observed value of $\Re$, say $\Re_o$. Conversely, with small values of $n_1, \ldots, n_g$ and N, a large value of $\Re_o$ may be associated with a relatively large value of $P(\delta \leq \delta_o | H_0)$.

2.2.2 Example Univariate MRPP Analysis with $v = 2$

Although multi-response permutation procedures were originally designed for analyzing multivariate response measurement scores, they can also be used for analyzing univariate data. Consider a comparison between two mutually exclusive groups of objects, S_1 and S_2, where a single response measurement, x, has been obtained from each object. For this example, there is $r = 1$ response measurement score for each object, $g = 2$ disjoint groups, and a total of $N = 6$ objects with $n_1 = 2$ and $n_2 = 4$ in treatment groups S_1 and S_2, respectively. Suppose that the $n_1 = 2$ observed response measurement scores for treatment group S_1 are $\{5,\ 4\}$ and the $n_2 = 4$ response measurement scores for treatment group S_2 are $\{2,\ 3,\ 7,\ 9\}$. The treatment-group sizes and the response measurement scores are deliberately kept small to simplify the example analysis. The treatment-group sizes and the univariate response measurement scores are listed in Fig. 2.3.

For this example analysis, let $v = 2, p = 2, r = 1$,

$$C_1 = \frac{n_1}{N} = \frac{2}{6}\ , \quad \text{and} \quad C_2 = \frac{n_2}{N} = \frac{4}{6}\ ,$$

Fig. 2.3 Example data with $g = 2$, $r = 1$, $n_1 = 2$, $n_2 = 4$, and $N = n_1 + n_2 = 6$

Group	Object	Value
S_1	ω_1	5
	ω_2	4
S_2	ω_3	2
	ω_4	3
	ω_5	7
	ω_6	9

so that the S_1 and S_2 treatment groups are weighted proportional to their group sizes of $n_1 = 2$ and $n_2 = 4$, respectively. For univariate response measurement scores with $r = 1$, Eq. (2.1) on p. 30 reduces to

$$\Delta(j,k) = \left(\left| x_j - x_k \right|^p \right)^{v/p} . \tag{2.7}$$

Thus, for treatment group S_1 with $n_1 = 2$ objects, $p = 2$, and $v = 2$, the generalized Minkowski distance function yields

$$\Delta(1,2) = \left(\left| 5 - 4 \right|^2 \right)^{2/2} = 1.00 ,$$

and for treatment group S_2 with $n = 4$ objects, the generalized Minkowski distance function yields

$$\Delta(3,4) = \left(\left| 2 - 3 \right|^2 \right)^{2/2} = 1.00 ,$$

$$\Delta(3,5) = \left(\left| 2 - 7 \right|^2 \right)^{2/2} = 25.00 ,$$

$$\Delta(3,6) = \left(\left| 2 - 9 \right|^2 \right)^{2/2} = 49.00 ,$$

$$\Delta(4,5) = \left(\left| 3 - 7 \right|^2 \right)^{2/2} = 16.00 ,$$

$$\Delta(4,6) = \left(\left| 3 - 9 \right|^2 \right)^{2/2} = 36.00 ,$$

and

$$\Delta(5,6) = \left(\left| 7 - 9 \right|^2 \right)^{2/2} = 4.00 .$$

Then following Eq. (2.3) on p. 31, the average distance-function values for all distinct pairs of objects in treatment groups S_i, $i = 1, 2$, are

$$\xi_1 = \binom{n_1}{2}^{-1} \Big[\Delta(1,2)\Big] = \binom{2}{2}^{-1} (1.00) = 1.00$$

and

$$\begin{aligned}\xi_2 &= \binom{n_2}{2}^{-1} \Big[\Delta(3,4) + \Delta(3,5) + \Delta(3,6) + \Delta(4,5) + \Delta(4,6) + \Delta(5,6)\Big] \\ &= \binom{4}{2}^{-1} (1.00 + 25.00 + 49.00 + 16.00 + 36.00 + 4.00) = 21.8333 \,.\end{aligned}$$

Following Eq. (2.2) on p. 31, the observed weighted mean of the ξ_1 and ξ_2 values, based on $v = 2$ and $C_i = n_i/N$ for $i = 1, 2$ is

$$\delta_\text{o} = C_1\xi_1 + C_2\xi_2 = \left(\frac{2}{6}\right)(1.00) + \left(\frac{4}{6}\right)(21.8333) = 14.8889 \,.$$

Smaller values of δ_o indicate a concentration of response measurement scores within the g treatment groups, whereas larger values of δ_o indicate a lack of concentration between response measurement scores among the g treatment groups [301]. The $N = 6$ objects can be partitioned into $g = 2$ treatment groups, S_1 and S_2, respectively, with $n_1 = 2$ and $n_2 = 4$ response measurement scores preserved in

$$M = \frac{N!}{n_1!\; n_2!} = \frac{6!}{2!\; 4!} = 15$$

possible, equally-likely ways. The $M = 15$ possible arrangements of the observed data in Fig. 2.3, along with the corresponding ξ_1, ξ_2, and δ values, are listed in Table 2.2 and ordered by the δ values from lowest to highest. The observed MRPP test statistic, $\delta_\text{o} = 14.8889$, obtained from the realized arrangement,

$$\{5, 4\} \quad \{2, 3, 7, 9\} \,,$$

(Order 9 in Table 2.2) is not unusual since five of the remaining δ values (δ_{11} to δ_{15}) exceed the observed value of $\delta_\text{o} = 14.8889$ and 10 values of δ (δ_1 to δ_{10}) are equal to or less than the observed value. If all arrangements of the $N = 6$ observed response measurement scores listed in Fig. 2.3 occur with equal chance, the exact probability value of $\delta_\text{o} = 14.8889$ computed on the $M = 15$ possible arrangements of the observed data with $n_1 = 2$ and $n_2 = 4$ response measurement scores preserved for

Table 2.2 Permutations of the observed data in Fig. 2.3 for treatment groups S_1 and S_2 with values for ξ_1, ξ_2, and δ based on $v = 2$, ordered by values of δ from lowest to highest

Order	S_1	S_2	ξ_1	ξ_2	δ
1	$\{7, 9\}$	$\{2, 5, 3, 4\}$	4.0000	3.3333	3.5556
2	$\{2, 3\}$	$\{5, 4, 7, 9\}$	1.0000	9.8333	6.8889
3	$\{2, 4\}$	$\{5, 3, 7, 9\}$	4.0000	13.3333	10.2222
4	$\{5, 9\}$	$\{2, 3, 4, 7\}$	16.0000	9.3333	11.5556
5	$\{3, 4\}$	$\{2, 5, 7, 9\}$	1.0000	17.8333	12.2222
6	$\{2, 5\}$	$\{3, 4, 7, 9\}$	9.0000	15.1667	13.1111
7	$\{5, 3\}$	$\{2, 4, 7, 9\}$	4.0000	19.3333	14.2222
8	$\{5, 7\}$	$\{2, 3, 4, 9\}$	4.0000	19.3333	14.2222
9	$\{5, 4\}$	$\{2, 3, 7, 9\}$	1.0000	21.8333	14.8889
10	$\{4, 9\}$	$\{2, 5, 3, 7\}$	25.0000	9.8333	14.8889
11	$\{4, 7\}$	$\{2, 5, 3, 9\}$	9.0000	19.1667	15.7778
12	$\{3, 7\}$	$\{2, 5, 4, 9\}$	16.0000	17.3333	16.8889
13	$\{2, 7\}$	$\{5, 3, 4, 9\}$	25.0000	13.8333	17.5556
14	$\{3, 9\}$	$\{2, 5, 4, 7\}$	36.0000	8.6667	17.7778
15	$\{2, 9\}$	$\{5, 3, 4, 7\}$	49.0000	5.8333	20.2222

each arrangement is

$$P\big(\delta \leq \delta_{\text{o}} | H_0\big) = \frac{\text{number of } \delta \text{ values} \leq \delta_{\text{o}}}{M} = \frac{10}{15} = 0.6667 \ .$$

For comparison, a conventional Student two-sample pooled t test calculated on the $N = 6$ response measurement scores listed in Fig. 2.3 yields an observed value of $t_{\text{o}} = -0.3004$. Assuming independence, normality, and homogeneity of variance, t is approximately distributed as Student's t under the null hypothesis with $N - 2 = 6 - 2 = 4$ degrees of freedom. Under the null hypothesis, the observed value of $t_{\text{o}} = -0.3004$ yields an approximate two-sided probability value of $P = 0.7789$.

Following Eq. (2.6) on p. 37, the exact average value of the $M = 15$ δ values listed in Table 2.2 is $\mu_\delta = 13.60$. Thus, the observed chance-corrected coefficient of agreement, following Eq. (2.5) on p. 37, is

$$\Re_{\text{o}} = 1 - \frac{\delta_{\text{o}}}{\mu_\delta} = 1 - \frac{14.8889}{13.60} = -0.0948 \ ,$$

indicating that within-group agreement is well below that expected by chance.

2.2.3 Example Univariate MRPP Analysis with $v = 1$

Permutation statistical tests and measures are data-dependent, distribution-free, and non-parametric; consequently, they require no distributional assumptions and make no estimates of population parameters. Thus, it is not necessary to set $v = 2$ and to square the response-measurement differences between objects. While conventional tests and measures that assume normality must estimate the mean and variance, μ_x and σ_x^2, of the normal distribution, both of which are based on squared deviations from the mean, permutation tests and measures do not assume normality and are not restricted to $v = 2$, which is not a metric distance function. A distance function based on $v = 1$ is an attractive alternative to $v = 2$ as it is a metric distance function, satisfies the triangle inequality, is robust to extreme values, provides an easy-to-understand ordinary Euclidean distance between objects, and ensures that the data and analysis spaces are congruent [284–287, 289, 295]. In addition, choosing $v = 1$ over $v = 2$ can make a substantial difference in the results of an MRPP analysis; see, for example, a discussion by Mielke and Berry in 2007 [297, pp. 45–50].

To illustrate the computation of δ with $v = 1$, consider the same finite sample of $N = 6$ objects listed in Fig. 2.3 on p. 40 and let S_1 and S_2 denote an exhaustive partitioning of the $N = 6$ objects into $g = 2$ disjoint treatment groups. As previously, let S_1 consist of $n_1 = 2$ objects, each with a single response measurement, and let S_2 consist of $n_2 = 4$ objects, each with a single response measurement.

Given the univariate data listed in Fig. 2.3, let $r = 1$, $p = 2$,

$$C_1 = \frac{n_1}{N} = \frac{2}{6} \,, \quad \text{and} \quad C_2 = \frac{n_2}{N} = \frac{4}{6} \,,$$

but in this case set $v = 1$ instead of $v = 2$, employing ordinary Euclidean distance instead of squared Euclidean distance between objects. Following Eq. (2.7) on p. 40 for treatment group S_1 with $n_1 = 2$ objects, $p = 2$, and $v = 1$, the generalized Minkowski distance function yields

$$\Delta(1,2) = \left(|(5-4|^2 \right)^{1/2} = 1.00 \,,$$

and for treatment group S_2 with $n = 4$ objects, the generalized Minkowski distance function yields

$$\Delta(3,4) = \left(|2-3|^2 \right)^{1/2} = 1.00 \,,$$

$$\Delta(3,5) = \left(|2-7|^2 \right)^{1/2} = 5.00 \,,$$

$$\Delta(3,6) = \left(|2-9|^2 \right)^{1/2} = 7.00 \,,$$

$$\Delta(4,5) = \left(|3-7|^2 \right)^{1/2} = 4.00 \,,$$

$$\Delta(4,6) = \left(|3-9|^2 \right)^{1/2} = 6.00 \,,$$

and

$$\Delta(5,6) = \left(\left|7-9\right|^2\right)^{1/2} = 2.00\ .$$

Then following Eq. (2.3) on p. 31, the average distance-function values for all distinct pairs of objects in treatment group S_i, $i = 1, 2$, are

$$\xi_1 = \binom{n_1}{2}^{-1} \left[\Delta(1,2)\right] = \binom{2}{2}^{-1} (1.00) = 1.00$$

and

$$\begin{aligned}\xi_2 &= \binom{n_2}{2}^{-1} \left[\Delta(3,4) + \Delta(3,5) + \Delta(3,6) + \Delta(4,5) + \Delta(4,6) + \Delta(5,6)\right] \\ &= \binom{4}{2}^{-1} (1.00 + 5.00 + 7.00 + 4.00 + 6.00 + 2.00) = 4.1667\ .\end{aligned}$$

Following Eq. (2.2) on p. 31, the observed weighted mean of the ξ_1 and ξ_2 values, based on $C_i = n_i/N$ for $i = 1, 2$ is

$$\delta_o = C_1\xi_1 + C_2\xi_2 = \left(\frac{2}{6}\right)(1.00) + \left(\frac{4}{6}\right)(4.1667) = 3.1111\ .$$

As in the previous MRPP example with $v = 2$, the $N = 6$ objects can be partitioned into $g = 2$ treatment groups, S_1 and S_2, with $n_1 = 2$ and $n_2 = 4$ response measurement scores preserved for each arrangement of the observed data in

$$M = \frac{N!}{n_1!\ n_2!} = \frac{6!}{2!\ 4!} = 15$$

possible, equally-likely ways. The $M = 15$ possible arrangements of the observed data in Fig. 2.3, along with the corresponding ξ_1, ξ_2, and δ values, are listed in Table 2.3 and ordered by the δ values from lowest to highest. The observed MRPP test statistic, $\delta_o = 3.1111$, obtained from the realized arrangement,

$$\{5, 4\} \quad \{2, 3, 7, 9\}\ ,$$

(Order 5 in Table 2.3) is not unusual since eight of the remaining δ values (δ_8 to δ_{15}) exceed the observed value of $\delta_o = 3.1111$ and seven values of δ (δ_1 to δ_7) are equal to or less than the observed value. If all arrangements of the $N = 6$ observed

Table 2.3 Permutations of the observed data in Fig. 2.3 for treatment groups S_1 and S_2 with values for ξ_1, ξ_2, and δ based on $v = 1$, ordered by values of δ from lowest to highest

Order	S_1	S_2	ξ_1	ξ_2	δ
1	{7, 9}	{2, 5, 3, 4}	2.0000	1.6667	1.7778
2	{2, 3}	{5, 4, 7, 9}	1.0000	2.8333	2.2222
3	{2, 4}	{5, 3, 7, 9}	2.0000	3.3333	2.8889
4	{3, 4}	{2, 5, 7, 9}	1.0000	3.8333	2.8889
5	{5, 4}	{2, 3, 7, 9}	1.0000	4.1667	3.1111
6	{5, 7}	{2, 3, 4, 9}	2.0000	3.6667	3.1111
7	{5, 9}	{2, 3, 4, 7}	4.0000	2.6667	3.1111
8	{2, 5}	{3, 4, 7, 9}	3.0000	3.5000	3.3333
9	{5, 3}	{2, 4, 7, 9}	2.0000	4.0000	3.3333
10	{4, 7}	{2, 5, 3, 9}	3.0000	3.8333	3.5556
11	{4, 9}	{2, 5, 3, 7}	5.0000	2.8333	3.5556
12	{2, 7}	{5, 3, 4, 9}	5.0000	3.1667	3.7778
13	{3, 7}	{2, 5, 4, 9}	4.0000	3.6667	3.7778
14	{2, 9}	{5, 3, 4, 7}	7.0000	2.1667	3.7778
15	{3, 9}	{2, 5, 4, 7}	6.0000	2.6667	3.7778

response measurement scores listed in Fig. 2.3 occur with equal chance, the exact probability value of $\delta_o = 3.1111$ computed on the $M = 15$ possible arrangements of the observed data with $n_1 = 2$ and $n_2 = 4$ response measurement scores preserved for each arrangement is

$$P\big(\delta \leq \delta_o | H_0\big) = \frac{\text{number of } \delta \text{ values} \leq \delta_o}{M} = \frac{7}{15} = 0.4667 \; .$$

For comparison, for the univariate data listed in Fig. 2.3 the exact probability value based on $v = 2$, $M = 15$, and $C_i = n_i/N$ for $i = 1, 2$ in the previous example is $P = 0.6667$. No comparison is made with the conventional Student two-sample t test as Student's t test is undefined for $v = 1$.

Following Eq. (2.6) on p. 37, the exact average value of the $M = 15$ δ values listed in Table 2.3 is $\mu_\delta = 3.20$. Thus, the observed chance-corrected coefficient of agreement, following Eq. (2.5) on p. 37, is

$$\Re_o = 1 - \frac{\delta_o}{\mu_\delta} = 1 - \frac{14.8889}{3.20} = +0.0278 \; ,$$

indicating very little within-group agreement above that expected by chance.

2.2.4 Example Bivariate MRPP Analysis with $v = 2$

In this second example, bivariate response measurement scores are used for simplicity to demonstrate a multivariate MRPP analysis. To illustrate the computation of MRPP with bivariate response measurement scores for each object, consider a finite

Fig. 2.4 Example data with $g = 2$, $r = 2$, $n_1 = 4$, $n_2 = 3$, and $N = n_1 + n_2 = 7$

		Values	
Group	Object	x_1	x_2
S_1	ω_1	5	1
S_1	ω_2	4	6
S_1	ω_3	5	2
S_1	ω_4	6	3
S_2	ω_5	2	3
S_2	ω_6	3	4
S_2	ω_7	2	4

sample of $N = 7$ objects and let S_1 and S_2 denote an exhaustive partitioning of the N objects into $g = 2$ disjoint treatment groups. Further, let S_1 consist of $n_1 = 4$ objects with $r = 2$ commensurate response measurement scores (x_{1i} and x_{2i}) on each object for $i = 1, \ldots, 4$, with $x_1' = (5,\ 1)$, $x_2' = (4,\ 6)$, $x_3' = (5,\ 2)$, and $x_4' = (6,\ 3)$, and let S_2 consist of $n_2 = 3$ objects with $r = 2$ commensurate response measurement scores (x_{1i} and x_{2i}) on each object for $i = 1, 2, 3$ with $x_5' = (2,\ 3)$, $x_6' = (3,\ 4)$, and $x_7' = (2,\ 4)$. The treatment group sizes and the response measurement scores are deliberately kept small to simplify the example analysis. The bivariate response measurement scores for the $N = 7$ objects are listed in Fig. 2.4.

For this example analysis, let $v = 2$, $p = 2$, $r = 2$,

$$C_1 = \frac{n_1}{N} = \frac{4}{7}\ , \quad \text{and} \quad C_2 = \frac{n_2}{N} = \frac{3}{7}\ ,$$

so that the S_1 and S_2 treatment groups are weighted proportional to their group sizes of $n_1 = 4$ and $n_2 = 3$, respectively. Following Eq. (2.1) on p. 30 for treatment group S_1 with $n_1 = 4$ objects, $p = 2$, and $v = 2$, the generalized Minkowski distance function yields

$$\Delta(1,2) = \left(|5-4|^2 + |1-6|^2\right)^{2/2} = 26.00\ ,$$

$$\Delta(1,3) = \left(|5-5|^2 + |1-2|^2\right)^{2/2} = 1.00\ ,$$

$$\Delta(1,4) = \left(|5-6|^2 + |1-3|^2\right)^{2/2} = 5.00\ ,$$

$$\Delta(2,3) = \left(|4-5|^2 + |6-2|^2\right)^{2/2} = 17.00\ ,$$

$$\Delta(2,4) = \left(|4-6|^2 + |6-3|^2\right)^{2/2} = 13.00\ ,$$

and

$$\Delta(3,4) = \left(|5-6|^2 + |2-3|^2\right)^{2/2} = \ 2.00\ ,$$

and for treatment group S_2 with $n_2 = 3$ objects, the generalized Minkowski distance function yields

$$\Delta(5,6) = \left(|2-3|^2 + |3-4|^2\right)^{2/2} = 2.00\ ,$$

$$\Delta(5,7) = \left(|2-2|^2 + |3-4|^2\right)^{2/2} = 1.00\ ,$$

and

$$\Delta(6,7) = \left(|3-2|^2 + |4-4|^2\right)^{2/2} = 1.00\ .$$

Then following Eq. (2.3) on p. 31, the average distance-function values for all distinct pairs of objects in treatment group S_i, $i = 1, 2$, are

$$\begin{aligned}\xi_1 &= \binom{n_1}{2}^{-1} \left[\Delta(1,2) + \Delta(1,3) + \Delta(1,4) + \Delta(2,3) + \Delta(2,4) + \Delta(3,4)\right] \\ &= \binom{4}{2}^{-1} (26.00 + 1.00 + 5.00 + 17.00 + 13.00 + 2.00) = 10.6667\end{aligned}$$

and

$$\begin{aligned}\xi_2 &= \binom{n_2}{2}^{-1} \left[\Delta(5,6) + \Delta(5,7) + \Delta(6,7)\right] \\ &= \binom{3}{2}^{-1} (2.00 + 1.00 + 1.00) = 1.3333\ .\end{aligned}$$

Following Eq. (2.2) on p. 31, the observed weighted mean of the ξ_1 and ξ_2 values, based on $v = 2$ and $C_i = n_i/N$ for $i = 1, 2$ is

$$\delta_\text{o} = C_1\xi_1 + C_2\xi_2 = \left(\frac{4}{7}\right)(10.6667) + \left(\frac{3}{7}\right)(1.3333) = 6.6667\ .$$

The $N = 7$ objects can be partitioned into $g = 2$ treatment groups, S_1 and S_2, with $n_1 = 4$ and $n_2 = 3$ response measurement scores preserved for each arrangement of the observed data in

$$M = \frac{N!}{n_1!\, n_2!} = \frac{7!}{4!\, 3!} = 35$$

possible, equally-likely ways. The $M = 35$ possible arrangements of the observed bivariate data in Fig. 2.4, along with the corresponding ξ_1, ξ_2, and δ values, are listed in Table 2.4 and ordered by the δ values from lowest to highest. The observed MRPP test statistic, $\delta_{\text{o}} = 6.6667$, obtained from the realized arrangement,

$$\{(5,1)(4,6)(5,2)(6,3)\} \quad \{(2,3)(3,4)(2,4)\} ,$$

(Order 3 in Table 2.4) is unusual since 32 of the remaining δ values (δ_4 to δ_{35}) exceed the observed value of $\delta_{\text{o}} = 6.6667$ and only two values of δ are less than the observed value: $\delta_1 = 4.0000$ and $\delta_2 = 6.4762$. If all arrangements of the $N = 7$ observed bivariate response measurement scores listed in Fig. 2.4 occur with equal chance, the exact probability value of $\delta_{\text{o}} = 6.6667$ computed on the $M = 35$ possible arrangements of the observed data with $n_1 = 4$ and $n_2 = 3$ response measurement scores preserved for each arrangement is

$$P\big(\delta \leq \delta_{\text{o}}|H_0\big) = \frac{\text{number of } \delta \text{ values} \leq \delta_{\text{o}}}{M} = \frac{3}{35} = 0.0857 .$$

A conventional Hotelling two-sample T^2 test is given by

$$T^2 = \frac{n_1 n_2}{N} \left(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2\right)' \mathbf{S}^{-1} \left(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2\right) , \tag{2.8}$$

where $\bar{\mathbf{y}}_1$ and $\bar{\mathbf{y}}_2$ denote vectors of mean differences between treatment groups S_1 and S_2, n_1 and n_2 are the number of interval-level multivariate response measurement scores in treatment groups S_1 and S_2, and $\mathbf{S}$ is a pooled variance–covariance matrix.

For the example data listed in Fig. 2.4, $\bar{y}_{11} = 5.00$, $s^2_{11} = 0.6167$, $\bar{y}_{12} = 3.00$, $s^2_{12} = 4.6667$, $\text{cov}(1,2)_1 = -1.00$, $\bar{y}_{21} = 2.3333$, $s^2_{21} = 0.3333$, $\bar{y}_{22} = 3.6667$, $s^2_{22} = 0.3333$, and $\text{cov}(1,2)_2 = +0.1667$. Then, $\bar{\mathbf{y}}_1 = \bar{y}_{11} - \bar{y}_{21} = 5.00 - 2.3333 = +2.6667$ and $\bar{\mathbf{y}}_2 = \bar{y}_{12} - \bar{y}_{22} = 3.00 - 3.6667 = -0.6667$.

The variance–covariance matrices for treatment groups S_1 and S_2 in Fig. 2.4 are

$$\hat{\boldsymbol{\Sigma}}_1 = \begin{bmatrix} 0.6667 & -1.0000 \\ -1.0000 & 4.6667 \end{bmatrix} \quad \text{and} \quad \hat{\boldsymbol{\Sigma}}_2 = \begin{bmatrix} 0.3333 & +0.1667 \\ +0.1667 & 0.3333 \end{bmatrix} ,$$

Table 2.4 Permutations of the observed data set in Fig. 2.4 for treatment groups S_1 and S_2 with values for ξ_1, ξ_2, and δ based on $v = 2$, ordered by values of δ from lowest to highest

Order	S_1	S_2	ξ_1	ξ_2	δ
1	{(4, 6) (2, 3) (3, 4) (2, 4)}	{(5, 1) (5, 2) (6, 3)}	5.0000	2.6667	4.0000
2	{(5, 1) (5, 2) (6, 3) (2, 3)}	{(4, 6) (3, 4) (2, 4)}	7.8333	4.6667	6.4762
3	{(5, 1) (4, 6) (5, 2) (6, 3)}	{(2, 3) (3, 4) (2, 4)}	10.6667	1.3333	6.6667
4	{(5, 1) (5, 2) (6, 3) (3, 4)}	{(4, 6) (2, 3) (2, 4)}	6.5000	7.3333	6.8571
5	{(5, 1) (5, 2) (6, 3) (2, 4)}	{(4, 6) (2, 3) (3, 4)}	9.3333	6.6667	8.1905
6	{(4, 6) (6, 3) (3, 4) (2, 4)}	{(5, 1) (5, 2) (2, 3)}	9.0000	8.0000	8.5714
7	{(5, 1) (2, 3) (3, 4) (2, 4)}	{(4, 6) (5, 2) (6, 3)}	8.0000	10.6667	9.1429
8	{(5, 1) (5, 2) (2, 3) (2, 4)}	{(4, 6) (6, 3) (3, 4)}	9.3333	9.3333	9.3333
9	{(5, 2) (2, 3) (3, 4) (2, 4)}	{(5, 1) (4, 6) (6, 3)}	5.8333	14.6667	9.6190
10	{(4, 6) (6, 3) (2, 3) (2, 4)}	{(5, 1) (5, 2) (3, 4)}	11.3333	7.3333	9.6190
11	{(4, 6) (5, 2) (6, 3) (3, 5)}	{(5, 1) (2, 3) (2, 4)}	9.1667	10.6667	9.8095
12	{(4, 6) (5, 2) (3, 4) (2, 4)}	{(5, 1) (6, 3) (2, 3)}	8.6667	11.3333	9.8095
13	{(5, 1) (5, 2) (2, 3) (3, 4)}	{(4, 6) (6, 3) (2, 4)}	7.8333	12.6667	9.9048
14	{(4, 6) (5, 2) (2, 3) (2, 4)}	{(5, 1) (6, 3) (3, 4)}	10.3333	9.3333	9.9048
15	{(4, 6) (6, 3) (2, 3) (3, 4)}	{(5, 1) (5, 2) (2, 4)}	9.8333	10.6667	10.1905
16	{(5, 1) (4, 6) (6, 3) (3, 4)}	{(5, 2) (2, 3) (2, 4)}	12.0000	8.0000	10.2857
17	{(5, 1) (4, 6) (2, 3) (2, 4)}	{(5, 2) (6, 3) (3, 4)}	13.1667	6.6667	10.3810
18	{(4, 6) (5, 2) (6, 3) (2, 4)}	{(5, 1) (2, 3) (3, 4)}	11.6667	9.3333	10.6667
19	{(6, 3) (2, 3) (3, 4) (2, 4)}	{(5, 1) (4, 6) (5, 2)}	7.8333	14.6667	10.7619
20	{(5, 1) (4, 6) (3, 4) (2, 4)}	{(5, 2) (6, 3) (2, 3)}	11.8333	9.3333	10.7619
21	{(5, 1) (6, 3) (2, 3) (2, 4)}	{(4, 6) (5, 2) (3, 4)}	11.6667	10.0000	10.9524
22	{(4, 6) (5, 2) (2, 3) (3, 4)}	{(5, 1) (6, 3) (2, 4)}	9.1667	13.3333	10.9524
23	{(5, 1) (6, 3) (2, 3) (3, 4)}	{(4, 6) (5, 2) (2, 4)}	9.8333	12.6667	11.0476
24	{(5, 1) (5, 2) (3, 4) (2, 4)}	{(2, 6) (6, 3) (2, 3)}	9.0000	14.0000	11.1429
25	{(5, 1) (4, 6) (6, 3) (2, 4)}	{(5, 2) (2, 3) (3, 4)}	14.5000	6.6667	11.1429
26	{(4, 6) (5, 2) (6, 3) (2, 3)}	{(5, 1) (3, 4) (2, 4)}	11.8333	10.6667	11.3333
27	{(5, 1) (4, 6) (6, 3) (2, 3)}	{(5, 2) (3, 4) (2, 4)}	14.3333	7.3333	11.3333
28	{(5, 1) (4, 6) (2, 3) (3, 4)}	{(5, 2) (6, 3) (2, 4)}	12.0000	10.6667	11.4286
29	{(5, 1) (4, 6) (5, 2) (3, 4)}	{(6, 3) (2, 3) (2, 4)}	11.6667	11.3333	11.5238
30	{(5, 1) (4, 6) (5, 2) (2, 3)}	{(6, 3) (3, 4) (2, 4)}	13.3333	9.3333	11.6190
31	{(5, 1) (6, 3) (3, 4) (2, 4)}	{(4, 6) (5, 2) (2, 3)}	10.6667	13.3333	11.8095
32	{(5, 2) (6, 3) (2, 3) (2, 4)}	{(5, 1) (4, 6) (3, 4)}	9.8333	14.6667	11.9048
33	{(5, 1) (4, 6) (5, 2) (2, 4)}	{(6, 3) (2, 3) (3, 4)}	13.8333	9.3333	11.9048
34	{(5, 2) (6, 3) (2, 3) (3, 4)}	{(5, 1) (4, 6) (2, 4)}	8.0000	17.3333	12.0000
35	{(5, 2) (6, 3) (3, 4) (2, 4)}	{(5, 1) (4, 6) (2, 3)}	8.5000	17.3333	12.2857

respectively, and the pooled variance–covariance matrix and its inverse are

$$\mathbf{S} = \begin{bmatrix} 0.5333 & -0.5333 \\ -0.5333 & 2.9333 \end{bmatrix} \quad \text{and} \quad \mathbf{S}^{-1} = \begin{bmatrix} +2.9167 & +0.4167 \\ +0.4167 & +0.4167 \end{bmatrix} ,$$

respectively.[7]

Following Eq. (2.8), the observed value of Hotelling's T^2 is

$$\begin{aligned} T_{\text{o}}^2 &= \frac{n_1 n_2}{N} (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)' \, \mathbf{S}^{-1} \, (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2) \\ &= \frac{(4)(3)}{7} \begin{bmatrix} +2.6667 & -0.6667 \end{bmatrix} \begin{bmatrix} +2.2917 & +0.4167 \\ +0.4167 & +0.4167 \end{bmatrix} \begin{bmatrix} +2.6667 \\ -0.6667 \end{bmatrix} \\ &= (1.7143)(15.00) = 25.7143 \end{aligned}$$

and the observed F-ratio for Hotelling's T^2 is

$$F_{\text{o}} = \frac{N-r-1}{r(N-r)} T_{\text{o}}^2 = \frac{7-2-1}{2(7-2)} (25.7145) = 10.2858 \; .$$

Assuming independence, normality, and homogeneity of variance, F is approximately distributed as Snedecor's F under the null hypothesis with $\nu_1 = r = 2$ and $\nu_2 = N - r - 1 = 7 - 2 - 1 = 4$ degrees of freedom. Under the null hypothesis, the observed value of $F_{\text{o}} = 10.2858$ yields an approximate probability value of $P = 0.0265$. While there is a considerable difference between the exact probability value of $P = 0.0857$ and the approximate probability value of $P = 0.0265$, it is not surprising, as Hotelling's T^2 test was not designed for samples as small as $n_1 = 4$ and $n_2 = 3$.

Following Eq. (2.6) on p. 37, the exact average value of the $M = 35$ δ values listed in Table 2.4 is $\mu_\delta = 10.0952$. Thus, the observed chance-corrected coefficient of agreement, following Eq. (2.5) on p. 37, is

$$\Re_{\text{o}} = 1 - \frac{\delta_{\text{o}}}{\mu_\delta} = 1 - \frac{6.6667}{10.0952} = +0.3396 \; ,$$

indicating approximately 34 % within-group agreement above that expected by chance.

---

[7] Each element of the $\mathbf{S}$ matrix is constructed from two corresponding elements in the $\hat{\boldsymbol{\Sigma}}$ matrices, weighted by the degrees of freedom, i.e., $n - 1$. For example, the first element of the $\mathbf{S}$ matrix is $0.5333 = [(4-1)(0.6667) + (3-1)(0.3333)]/(4+3-2)$.

Fig. 2.5 Example data with $g = 2$, $r = 2$, $n_1 = 4$, $n_2 = 3$, and $N = n_1 + n_2 = 7$

		Values	
Group	Object	x_1	x_2
S_1	ω_1	5	1
S_1	ω_2	4	6
S_1	ω_3	5	2
S_1	ω_4	6	3
S_2	ω_5	2	3
S_2	ω_6	3	4
S_2	ω_7	2	4

2.2.5 Example Bivariate MRPP Analysis with $v = 1$

As mentioned in the univariate example on p. 43, the choice of v can make a substantial difference in the results of an MRPP analysis. To illustrate the computation of MRPP with bivariate data and $v = 1$, consider the same finite sample of $N = 7$ objects listed in Fig. 2.4 on p. 46 and let S_1 and S_2 denote an exhaustive partitioning of the N objects into $g = 2$ disjoint treatment groups. As previously, let S_1 consist of $n_1 = 4$ objects with $r = 2$ commensurate response measurement scores (x_{1i} and x_{2i}) on each object for $i = 1, \ldots, 4$, with $x_1' = (5,\ 1)$, $x_2' = (4,\ 6)$, $x_3' = (5,\ 2)$, and $x_4' = (6,\ 3)$, and let S_2 consist of $n_2 = 3$ objects with $r = 2$ commensurate response measurement scores (x_{1i} and x_{2i}) on each object for $i = 1, 2, 3$ with $x_5' = (2,\ 3)$, $x_6' = (3,\ 4)$, and $x_7' = (2,\ 4)$.

The bivariate response measurement scores for the $N = 7$ objects are listed in Fig. 2.4 on p. 46 and are replicated in Fig. 2.5 for convenience.

For this example analysis, let $r = 2$, $C_1 = n_1/N = 4/7$, $C_2 = n_2/N = 3/7$, and $p = 2$, but in this case set $v = 1$ instead of $v = 2$, employing ordinary Euclidean distance between objects. Following Eq. (2.1) on p. 30 for treatment group S_1 with $n_1 = 4$ objects, $p = 2$, and $v = 1$, the generalized Minkowski distance function yields

$$\Delta(1,2) = \left(|5-4|^2 + |1-6|^2\right)^{1/2} = 5.0990\ ,$$

$$\Delta(1,3) = \left(|5-5|^2 + |1-2|^2\right)^{1/2} = 1.0000\ ,$$

$$\Delta(1,4) = \left(|5-6|^2 + |1-3|^2\right)^{1/2} = 2.2361\ ,$$

$$\Delta(2,3) = \left(|4-5|^2 + |6-2|^2\right)^{1/2} = 4.1231\ ,$$

$$\Delta(2,4) = \left(|4-6|^2 + |6-3|^2\right)^{1/2} = 3.6056\ ,$$

and

$$\Delta(3,4) = \left(\left|5-6\right|^2 + \left|2-3\right|^2\right)^{1/2} = 1.4142\ ,$$

and for treatment group S_2 with $n_2 = 3$ objects, the generalized Minkowski distance function yields

$$\Delta(5,6) = \left(\left|2-3\right|^2 + \left|3-4\right|^2\right)^{1/2} = 1.4142\ ,$$

$$\Delta(5,7) = \left(\left|2-2\right|^2 + \left|3-4\right|^2\right)^{1/2} = 1.0000\ ,$$

and

$$\Delta(6,7) = \left(\left|3-2\right|^2 + \left|4-4\right|^2\right)^{1/2} = 1.0000\ .$$

Then, following Eq. (2.3) on p. 31, the average distance-function values for all distinct pairs of objects in treatment group S_i, $i = 1, 2$, are

$$\begin{aligned}\xi_1 &= \binom{n_1}{2}^{-1}\Big[\Delta(1,2) + \Delta(1,3) + \Delta(1,4) + \Delta(2,3) + \Delta(2,4) + \Delta(3,4)\Big] \\ &= \binom{4}{2}^{-1}(5.0990 + 1.0000 + 2.2361 + 4.1231 + 3.6056 + 1.4142) \\ &= 2.9130\end{aligned}$$

and

$$\begin{aligned}\xi_2 &= \binom{n_2}{2}^{-1}\Big[\Delta(5,6) + \Delta(5,7) + \Delta(6,7)\Big] \\ &= \binom{3}{2}^{-1}(1.4142 + 1.0000 + 1.0000) = 1.1381\ .\end{aligned}$$

Following Eq. (2.2) on p. 31, the observed weighted mean of the ξ_1 and ξ_2 values, based on $v = 1$ and $C_i = n_i/N$ for $i = 1, 2$ is

$$\delta_{\text{o}} = C_1\xi_1 + C_2\xi_2 = \left(\frac{4}{7}\right)(2.9130) + \left(\frac{3}{7}\right)(1.1381) = 2.1523\ .$$

The $N = 7$ objects listed in Fig. 2.5 can be partitioned into $g = 2$ treatment groups, S_1 and S_2, with $n_1 = 4$ and $n_2 = 3$ response measurement scores preserved for each arrangement of the observed data in

$$M = \frac{N!}{n_1!\; n_2!} = \frac{7!}{4!\; 3!} = 35$$

possible, equally-likely ways. The $M = 35$ possible arrangements of the observed data in Fig. 2.5, along with the corresponding ξ_1, ξ_2, and δ values, are listed in Table 2.5 and ordered by the δ values from lowest to highest. The observed MRPP test statistic, $\delta_{\text{o}} = 2.1523$, obtained from the realized arrangement,

$$\{(5,1)(4,6)(5,2)(6,3)\} \quad \{(2,3)(3,4)(2,4)\}\;,$$

(Order 2 in Table 2.5) is unusual since 33 of the remaining δ values (δ_3 to δ_{35}) exceed the observed value of $\delta_{\text{o}} = 2.1523$ and only one value is less than the observed value: $\delta_1 = 1.8152$. If all arrangements of the $N = 7$ observed bivariate response measurement scores listed in Fig. 2.5 occur with equal chance, the exact probability value of $\delta_{\text{o}} = 2.1523$ computed on the $M = 35$ possible arrangements of the observed data with $n_1 = 4$ and $n_2 = 3$ response measurement scores preserved for each arrangement is

$$P\big(\delta \leq \delta_{\text{o}}|H_0\big) = \frac{\text{number of } \delta \text{ values} \leq \delta_{\text{o}}}{M} = \frac{2}{35} = 0.0571\;.$$

For comparison, for the bivariate response measurement scores listed in Fig. 2.5 the exact probability value based on $v = 2$ and $C_i = n_i/N$ for $i = 1, 2$ in the first example is $P = 0.0857$. No comparison is made with the conventional Hotelling T^2 test as Hotelling's T^2 is undefined for $v = 1$.

Following Eq. (2.6) on p. 37, the exact average value of the $M = 35$ δ values listed in Table 2.5 is $\mu_\delta = 2.9475$. Thus, the observed chance-corrected coefficient of agreement, following Eq. (2.5) on p. 37, is

$$\Re_{\text{o}} = 1 - \frac{\delta_{\text{o}}}{\mu_\delta} = 1 - \frac{2.1523}{2.9475} = +0.2698\;,$$

indicating approximately 27 % within-group agreement above that expected by chance.

2.3 Coda

Chapter 2 provided the foundation for Multi-Response Permutation Procedures (MRPP), with special emphasis on the generalized Minkowski distance function, $\Delta(x, y)$, as defined in Eq. (2.1) on p. 30; δ, the weighted mean of the specified distance function values for all distinct pairs of objects in treatment group S_i for

Table 2.5 Permutations of the observed data set in Fig. 2.5 for treatment groups S_1 and S_2 with values for ξ_1, ξ_2, and δ based on $v = 1$, ordered by values of δ from lowest to highest

Order	S_1	S_2	ξ_1	ξ_2	δ
1	{(4, 6) (2, 3) (3, 4) (2, 4)}	{(5, 1) (5, 2) (6, 3)}	2.0140	1.5501	1.8152
2	{(5, 1) (4, 6) (5, 2) (6, 3)}	{(2, 3) (3, 4) (2, 4)}	2.9130	1.1381	2.1523
3	{(5, 1) (5, 2) (6, 3) (2, 3)}	{(4, 6) (3, 4) (2, 4)}	2.5697	2.0215	2.3347
4	{(5, 1) (5, 2) (6, 3) (3, 4)}	{(4, 6) (2, 3) (2, 4)}	2.3744	2.4780	2.4188
5	{(5, 1) (5, 2) (6, 3) (2, 4)}	{(4, 6) (2, 3) (3, 4)}	2.7703	2.4186	2.6196
6	{(5, 1) (2, 3) (3, 4) (2, 4)}	{(4, 6) (5, 2) (6, 3)}	2.4780	3.0476	2.7221
7	{(4, 6) (6, 3) (3, 4) (2, 4)}	{(5, 1) (5, 2) (2, 3)}	2.8259	2.5893	2.7245
8	{(5, 2) (2, 3) (3, 4) (2, 4)}	{(5, 1) (4, 6) (6, 3)}	2.1684	3.6469	2.8020
9	{(6, 3) (2, 3) (3, 4) (2, 4)}	{(5, 1) (4, 6) (5, 2)}	2.4499	3.4074	2.8603
10	{(5, 1) (5, 2) (2, 3) (2, 4)}	{(4, 6) (6, 3) (3, 4)}	2.7693	3.0013	2.8687
11	{(4, 6) (6, 3) (2, 3) (2, 4)}	{(5, 1) (5, 2) (3, 4)}	3.1938	2.4780	2.8870
12	{(4, 6) (5, 2) (6, 5) (3, 4)}	{(5, 1) (2, 3) (2, 4)}	2.8949	2.9494	2.9183
13	{(4, 6) (6, 3) (2, 3) (3, 4)}	{(5, 1) (5, 2) (2, 4)}	3.0039	2.9494	2.9806
14	{(4, 6) (5, 2) (3, 4) (2, 4)}	{(5, 1) (6, 3) (2, 3)}	2.7703	3.2805	2.9890
15	{(5, 1) (5, 2) (2, 3) (3, 4)}	{(4, 6) (6, 3) (2, 4)}	2.6027	3.5190	2.9954
16	{(5, 1) (4, 6) (2, 3) (2, 4)}	{(5, 2) (6, 3) (3, 4)}	3.3969	2.4683	2.9989
17	{(5, 1) (4, 6) (6, 3) (3, 4)}	{(5, 2) (2, 3) (2, 4)}	3.3241	2.5893	3.0092
18	{(4, 6) (5, 2) (2, 3) (2, 4)}	{(5, 1) (6, 3) (3, 4)}	3.0542	3.0013	3.0315
19	{(5, 1) (4, 6) (3, 4) (2, 4)}	{(5, 2) (6, 3) (2, 3)}	3.1686	2.8588	3.0359
20	{(5, 1) (4, 6) (5, 2) (3, 4)}	{(6, 3) (2, 3) (2, 4)}	3.1487	3.0410	3.1026
21	{(4, 6) (5, 2) (6, 3) (2, 4)}	{(5, 1) (2, 3) (3, 4)}	3.2833	2.8751	3.1084
22	{(5, 1) (6, 3) (2, 3) (2, 4)}	{(4, 6) (5, 2) (3, 4)}	3.2012	3.0625	3.1418
23	{(5, 1) (4, 6) (5, 2) (2, 3)}	{(6, 3) (3, 4) (2, 4)}	3.4326	2.7618	3.1451
24	{(5, 1) (5, 2) (3, 4) (2, 4)}	{(2, 6) (6, 3) (2, 3)}	2.7137	3.7370	3.1523
25	{(4, 6) (5, 2) (6, 3) (2, 3)}	{(5, 1) (3, 4) (2, 4)}	3.3184	2.9494	3.1603
26	{(5, 1) (4, 6) (6, 3) (2, 4)}	{(5, 2) (2, 3) (3, 4)}	3.6891	2.4683	3.1659
27	{(4, 6) (5, 2) (2, 3) (3, 4)}	{(5, 1) (6, 3) (2, 4)}	2.8949	3.5339	3.1688
28	{(5, 1) (4, 6) (2, 3) (3, 4)}	{(5, 2) (6, 3) (2, 4)}	3.2610	3.0476	3.1695
29	{(5, 1) (4, 6) (6, 3) (2, 3)}	{(5, 2) (3, 4) (2, 4)}	3.6920	2.4780	3.1717
30	{(5, 2) (6, 3) (2, 3) (2, 4)}	{(5, 1) (4, 6) (3, 4)}	2.8842	3.6469	3.2111
31	{(5, 1) (4, 6) (5, 2) (2, 4)}	{(6, 3) (2, 3) (3, 4)}	3.4831	2.8588	3.2156
32	{(5, 1) (6, 3) (2, 3) (3, 4)}	{(4, 6) (5, 2) (2, 4)}	3.0039	3.5190	3.2247
33	{(5, 2) (6, 3) (2, 3) (3, 4)}	{(5, 1) (4, 6) (2, 4)}	2.6636	4.0567	3.2606
34	{(5, 2) (6, 3) (3, 4) (2, 4)}	{(5, 1) (4, 6) (2, 3)}	2.6889	4.1034	3.2951
35	{(5, 1) (6, 3) (3, 4) (2, 4)}	{(4, 6) (5, 2) (2, 3)}	3.0616	3.6303	3.3053

$i = 1, \ldots, g$, as defined in Eq. (2.2) on p. 31; and $\mathfrak{R}$, the chance-corrected within-group coefficient of agreement, as defined in Eq. (2.4) on p. 33. Chapters 3 and 4 provide applications of MRPP for completely randomized data at the interval level of measurement, Chaps. 5 and 6 provide applications of MRPP for completely randomized data at the ordinal (ranked) level of measurement, and Chap. 7 provides applications of MRPP for completely randomized data at the nominal (categorical) level of measurement.

Chapter 3

Chapter 3 establishes the relationship between the MRPP test statistics, δ and $\Re$, and selected conventional tests and measures designed for the analysis of completely randomized data at the interval level of measurement. Considered in Chap. 3 are Student's two-sample t test with interval-level univariate response measurement scores, Hotelling's two-sample T^2 test with interval-level multivariate response measurement scores, one-way fixed-effects analysis of variance (ANOVA) with interval-level univariate response measurement scores, and one-way multivariate analysis of variance (MANOVA) with interval-level multivariate response measurement scores.