Permutation Statistical Methods: Calculation Efficiencies

Kenneth J Berry[1]*, Janis E Johnston[1], Paul W Mielke Jr[2], Howard W Mielke[3] and Lindsay A Johnston[4]

[1]Department of Sociology, Colorado State University, USA

[2]Department of Statistics, Colorado State University, Fort Collins, USA

[3]Department of Pharmacology, Tulane University, USA

[4]US Air Force, JBER-Elmendorf Family Health Clinic, USA

Submission: October 17, 2017; **Published:** January 22, 2018

*Corresponding author: Kenneth J Berry, Department of Sociology, Colorado State University, Fort Collins, USA, Email: berry@mail.colostate.edu

Abstract

Permutation statistical methods have much to recommend them as they are distribution-free, appropriate for non-random samples, and provide exact probability values. However, permutation statistical methods, by their very design, are computationally intensive. Eight calculation efficiencies for permutation statistical methods are presented and illustrated with example data sets.

Keywords : Exact methods; Permutation methods; Probability; Mathematical recursion; Resampling methods

Introduction

Permutation statistical methods were initially developed by RA Fisher [1], RC Geary [2], T Eden & F Yates [3], EJG Pitman [4-6] and other mathematicians and scientists in the 1920s and 1930s for validating the normality and homogeneity assumptions of classical statistical methods. Subsequently, permutation statistical methods have emerged as an approach to data analysis in their own right. Permutation statistical methods possess several advantages over classical statistical methods. First, permutation tests are entirely data-dependent in that all the information required for analysis is contained within the observed data. Second, permutation tests are appropriate for non-random samples, such as are common in many fields of research. Third, permutation tests are distribution-free in that they do not depend on the assumptions associated with traditional parametric tests. Fourth, permutation tests provide exact probability values based on the discrete permutation distribution of equally-likely test statistic values. Fifth, permutation tests are ideal for small data sets.

Two types of permutation tests are common in the statistical literature: exact and Monte Carlo resampling tests. Although the two types of permutation statistical tests are methodologically quite different, both types are based on the same specified null hypothesis. In an exact permutation statistical test, the first step is to calculate a test statistic value for the observed data. Second,

all possible, equally-likely arrangements of the observed data are enumerated. Third, the desired test statistic is calculated for each arrangement of the observed data. The probability of obtaining the observed value of the test statistic, or one more extreme, is the proportion of the enumerated test statistics with values equal to or more extreme than the value of the observed test statistic. For large samples the total number of possible arrangements can be considerable and exact permutation methods are quickly rendered impractical. For example, permuting two small samples of sizes.

$n_1 = n_2 = 30$ yields

$$M = \frac{(n_1 + n_2)!}{n_1! \, n_2!} = \frac{(30+30)!}{30! \, 30!} = 118,264,581,564,861,424 \qquad (1)$$

Arrangement of the observed data set far too many statistical values to compute in a reasonable amount of time.

When exact permutation procedures become intractable, a random subset of all possible arrangements of the observed data can be analyzed, providing approximate, but highly accurate, probability values. Resampling permutation tests generate and examine a Monte Carlo random subset of all possible, equally-likely arrangements of the observed response measurements. For each randomly selected arrangement of the observed data, the desired test statistic is calculated. The probability of obtaining the observed value of the test statistic, or one more extreme,

is the proportion of the randomly selected test statistics with values equal to or more extreme than the value of the observed test statistic. With a sufficient number of random samples, a probability value can be computed to any reasonable accuracy. The current recommended practice is to use randomly selected arrangements of the observed data to ensure a probability value $L = 1,000,000$ with three decimal places of accuracy [7].

While permutation statistical tests do not require random sampling, normality, homogeneity of variance, or large sample sizes, and are also completely data-dependent, a potential drawback is the amount of computation required, with exact permutation tests formerly being unrealistic for many statistical analyses. Even Monte Carlo resampling permutation tests often require the enumeration of millions of random arrangements of the observed data in order to guarantee sufficient accuracy. For many years, exact tests were considered to be impractical, but modern computers make it possible to generate hundreds of millions of permutations in just a few minutes. In addition, Monte Carlo permutation methods can be inefficient due to millions of calls to a pseudorandom number generator (PRNG).

Eight innovations mitigate this problem. Seven of the eight innovations pertain to both exact and Monte Carlo resampling permutation methods, while the eighth innovation is specific to Monte Carlo resampling methods. First, high-speed computing makes possible exact permutation statistical tests in which all possible arrangements of the observed data are generated and examined. Second, examination of all combinations of arrangements of the observed data, instead of all permutations, yields the identical exact probability values with considerable savings in computation time. Third, mathematical recursion with an arbitrary initial value greatly simplifies difficult computations, such as large factorial expressions. Fourth, calculation of only the variable components of the selected test statistic greatly simplifies calculation. Fifth, for certain classes of problems with a limited number of discrete values it is possible to substantially reduce the number of permutation necessary to accomplish an exact test. Sixth, for matched-pairs and other block designs the required number of arrangements of the observed data can be greatly reduced by holding one set of blocks constant. Seventh, the utilization of partition theory for goodness-of-fit tests can reduce the required number of permutations of the observed data. Eighth, Monte Carlo resampling methods can provide highly-accurate approximate probability values for samples of any size with substantial reduction in calculation. Combinations of any or all of the eight calculation efficiencies provide highly accurate permutation analyses that would otherwise be prohibitive.

High-speed computing

As Berry, Johnston, and Mielke observed in 2014, one has only to observe the hordes of the digitally distracted trying to navigate a crowded sidewalk with their various smart-phones, pads, pods, and tablets to realize that computing power, speed, and accessibility have finally arrived [8]. As Martin Hilbert documented in 2012, just one percent of the world's capacity to store information was in digital format, but by year 2000 digital represented 25 percent of the total world's memory [9]. The year 2002 marked the start of the digital age, as 2002 was the year that humankind first stored more information in digital than in analog form. By 2007 over 97 percent of the world's storage capacity was digital. Moreover, it was estimated in 2012 that ninety percent of the data stored in the world had been created in just the previous two years. Prior to 2001, data storage was measured in bytes, kilobytes (10^3), and occasionally in megabytes (10^6); now data storage is measured in gigabytes (10^9), terabytes (10^{12}), petabytes (10^{15}), exabytes (10^{18}), zettabytes (10^{21}), and even yottabytes (10^{24}).

In 2000, the Intel Pentium processor contained 42 million transistors and ran at 1.5GHz. In the spring of 2010, Intel released the Itanium processor, code-named Tukwila after a town in the state of Washington, containing 1.4 billion transistors and running at 2.53GHz. On 4 June 2013 Intel announced the Haswell processor, named after a small town of 65 people in southeastern Colorado with 1.4 billion 3-D chips and running at 3.50GHz. The latest generation of Haswell processors, the i7-4790 processor, currently executes at 4.00GHz with turbo-boost to 4.40GHz. In April of 2017, Intel introduced the Optane memory module which, when coupled with a seventh generation Intel Core-based system, has the potential to increase desktop computer performance by 28 percent.

Permutation methods

While not widely available to researchers, by 2010 mainframe computers were measuring computing speeds in teraflops. To emphasize the progress of computing, in 1951 the Remington Rand Corporation introduced the UNIVAC computer running at 1,905 flops, which with ten mercury delay line memory tanks could store 20,000 bytes of information; in 2008 the IBM Corporation supercomputer, code-named Roadrunner, reached a sustained performance of one petaflops -a quadrillion operations per second; in 2010 the Cray Jaguar was named the world's fastest computer performing at a sustained speed of 1.75 petaflops with 360 terabytes of memory; and in November of 2010 China exceeded the computing speed of the Cray Jaguar by 57 percent with the introduction of China's Tianhe-1A supercomputer performing at 2.67 petaflops.

In October of 2011, China broke the petaflops barrier again with the introduction of the Sunway Blue light MPP. In late 2011, the IBM Yellowstone supercomputer was installed at the National Center for Atmospheric Research (NCAR) Wyoming Supercomputer Center in Cheyenne, Wyoming. After months of testing, the Wyoming Supercomputer Center officially opened on Monday, 15 October 2012. Yellowstone was a 1.6 petaflops machine with 149.2 terabytes of memory and 74,592 processor cores and replaced an IBM Bluefire supercomputer installed in 2008 that had a peak speed of 76 teraflops. Also in late 2011, IBM

unveiled the Blue Gene\P and \Q supercomputing processing systems that can achieve 20 petaflops. At the same time, IBM filed a patent for a massive supercomputing system capable of 107 petaflops.

From a more general perspective, in 1977 the Tandy Corporation released the TRS-80, the first fully assembled personal computer, distributed through Radio Shack stores. The TRS-80 had 4MB of RAM and ran at 1.78MHz. By way of comparison, in 2010 the Apple iPhone had 131,072 times the memory of the TRS-80 and was about 2,000 times faster, running at one GHz. In 2012, Sequoia, an IBM Blue Gene/Q supercomputer was installed at Lawrence Livermore National Laboratory (LLNL) in Livermore, California. In June of 2012 Sequoia officially became the most powerful supercomputer in the world. Sequoia is capable of 16.32 petaflops-more than 16 quadrillion calculations a second-which is 55 percent faster than Japan's K supercomputer, ranked number 2, and more than five times faster than China's Tianhe-1A, which was the fastest supercomputer in the world in 2010.

Finally, high-speed computers have dramatically changed the field of computational statistics. The future of high-speed computing appears very promising for exact and Monte Carlo resampling permutation statistical methods. Combined with other calculation efficiencies, it can safely be said that permutation methods have the potential to provide exact or resampling probability values in an efficient manner for a wide variety of statistical applications.

Analysis with combinations

Although permutation statistical methods are known by the attribution "permutation," they are, in fact, not based on all possible permutations of the observed data. Instead exact permutation methods are typically based on all possible combinations of the observed data. Conversely, a combination lock is not based on combinations of numbers or letters, but is instead based on all possible permutations of the numbers or letters. A simple example will illustrate. Consider N = 8 objects that are to be divided into two groups A and B, where $n_A = n_B = 4$. The purpose is to compare differences between the two groups, such as a mean or median difference. Let the eight objects be designated $\{a, b, c, d, e, f, g, h\}$. For group A, the first object can be chosen in eight different ways, the second object in seven ways, the third in six ways, and the fourth object in five ways. Once these four members of Group A are chosen, the membership of Group B is fixed, since the remaining four objects are assigned to Group B.

Of the 8×7×6×5 =1,680 ways in which the four objects can be arranged for Group A, each individual quartet of objects will appear in a series of permutations. Thus, the quartet $\{a, b, c, d\}$ can be permuted as $\{a, b, d, c\}$, $\{b, a, c, d\}$, $\{c, d, b, a\}$, and so on. The number of different permutations for a group of four different objects is $4! = 4 \times 3 \times 2 \times 1 = 24$. Thus, each distinctive quartet will appear in 24 ways among the 1,680 possible arrangements. Therefore, 1,680

divided by 24 yields 70 distinctive quartets that could be formed by dividing eight objects into two groups of four objects each. The number of quartets can conveniently be expressed as

$$\frac{(n_A + n_B)!}{n_A! n_B!} = \frac{(4+4)!}{4!4!} = \frac{40,320}{576} = 70. \quad (2)$$

Now, half of these arrangements are similar but opposite. Thus the quartet $\{a, b, c, d\}$ might be in Group A and the quartet $\{e, f, g, h\}$ might be in Group B, or vice versa, yielding the same absolute difference. Consequently, there are really only 70/2 = 35 distinctly different pairs of quartets to be considered. The 35 possible arrangements for objects $\{a, b, c, d, e, f, g, h\}$ are listed in Table 1 in Gray-code order. The next (36th) possible arrangement would be $\{a, b, c, h\}$ in Group A and $\{d, e, f, g\}$ in Group B, which is simply the reverse of arrangement 35, i.e., $\{d, e, f, g\}$ in Group A and $\{a, b, c, h\}$ in Group B, yielding the same absolute mean or median difference. A substantial amount of calculation can be eliminated by considering all possible combinations instead of all possible permutations, with no loss of accuracy. In this case, a decrease from 1,680 to 35 arrangements to be considered, a reduction of 98% (Table 1).

Table 1: Listing of the 35 arrangements of objects {a, b, c, d, e, f, g, h} into two groups of four objects each.

Number	Group A	Group B	Number	Group A	Group B
1	a, b, c, d	e, f, g, h	19	b, c, d, g	a, e, f, h
2	a, b, c, e	d, f, g, h	20	a, b, e, g	c, d, f, h
3	a, b, d, e	c, f, g, h	21	a, c, e, g	b, d, f, h
4	a, c, d, e	b, f, g, h	22	b, c, e, g	a, d, f, h
5	b, c, d, e	a, f, g, h	23	a, d, e, g	b, c, f, h
6	a, b, c, f	d, e, g, h	24	b, d, e, g	a, c, f, h
7	a, b, d, f	c, e, g, h	25	c, d, e, g	a, b, f, h
8	a, c, d, f	b, e, g, h	26	a, b, f, g	c, d, e, h
9	b, c, d, f	a, e, g, h	27	a, c, f, g	b, d, e, h
10	a, b, e, f	c, d, g, h	28	b, c, f, g	a, d, e, h
11	a, c, e, f	b, d, g, h	29	a, d, f, g	b, c, e, h
12	b, c, e, f	a, d, g, h	30	b, d, f, g	a, c, e, h
13	a, d, e, f	b, c, g, h	31	c, d, f, g	a, b, e, h
14	b, d, e, f	a, c, g, h	32	a, e, f, g	b, c, d, h
15	c, d, e, f	a, b, g, h	33	b, e, f, g	a, c, d, h
16	a, b, c, g	d, e, f, h	34	c, e, f, g	a, b, d, h
17	a, b, d, g	c, e, f, h	35	d, e, f, g	a, b, c, h
18	a, c, d, g	b, e, f, h			

Example analysis: Consider a sample of $N = 8$ objects with values $\{38, 39, 40, 43, 48, 49, 52, 57\}$. The $N = 8$ objects are divided into two groups, A and B with $n_A = 4$ objects in Group A and $n_B = 4$ objects in Group B. The objects in Group A have values of $\{43, 49, 52, 57\}$ and the objects in Group B have values of $\{38, 39, 40, 48\}$, yielding means $\bar{x}_A = 50.25$ and $\bar{x}_B = 41.25$, and a mean deference of $\bar{x}_A - \bar{x}_B = 50.25 - 41.25 = +9.00$. Now consider the data from a permutation perspective. Table 2 lists the 35 possible arrangements of the N = 8 values with $n_A = n_B = 4$ preserved for each arrangement, the mean values,

$\overline{x}_A$ and $\overline{x}_B$, and the 35 mean differences, $\overline{x}_A - \overline{x}_B$. Inspection of Table 2 reveals that a mean difference of $\overline{x}_A - \overline{x}_B = 50.25 - 41.25 = +9.00$ or greater in favor of Group A occurs only twice in the 35 possible arrangements, i.e., row 1 with mean difference +11.50 and row 2 with mean difference +9.00. If Table 2 were to be completed to form all 70 arrangements, a mean difference of 9.00 or greater would also occur twice. Thus, for a two-sided test the exact probability is P = 4/70 = 0.0571, and for a one-sided test the exact probability is P = 2/70 = 0.0286.

Table 2: Listing of the 35 arrangements of objects {a, b, c, d, e, f, g, h} into two groups of four objects each.

Number	Group A	Group B	$\overline{x}_A$	$\overline{x}_B$	$\overline{x}_A - \overline{x}_B$
1	48, 49, 52, 57	38, 39, 40, 43	51.5	40	+11.50
2	43, 49, 52, 57	38, 39, 40, 48	50.25	41.25	+9.00
3	43, 48, 52, 57	38, 39, 40, 49	50	41.5	+8.50
4	40, 49, 52, 57	38, 39, 43, 48	49.5	42	+7.50
5	39, 49, 52, 57	38, 40, 43, 48	49.25	42.25	+7.00
6	40, 48, 52, 57	38, 39, 43, 49	49.25	42.25	+7.00
7	43, 48, 49, 57	38, 39, 40, 52	49.25	42.25	+7.00
8	38, 49, 52, 57	39, 40, 43, 48	49.99	42.5	+6.50
9	39, 48, 52, 57	38, 40, 43, 49	49	42.5	+6.50
10	38, 48, 52, 57	39, 40, 43, 49	48.75	42.75	+6.00
11	40, 48, 49, 57	38, 39, 43, 52	48.5	43	+5.50
12	39, 48, 49, 57	38, 40, 43, 52	48.25	43.25	+5.00
13	40, 43, 52, 57	38, 39, 48, 49	48	43.5	+4.50
14	38, 48, 49, 57	39, 40, 43, 52	48	43.5	+4.50
15	39, 43, 52, 57	38, 40, 48, 49	47.75	43.75	+4.00
16	38, 43, 52, 57	39, 40, 48, 49	47.5	44	+3.50
17	40, 43, 49, 57	38, 39, 48, 52	47.25	44.25	+3.00
18	39, 40, 52, 57	38, 43, 48, 49	47	44.5	+2.50
19	39, 43, 49, 57	38, 40, 48, 52	47	44.5	+2.50
20	40, 43, 48, 57	38, 39, 49, 52	47	44.5	+2.50
21	39, 43, 48, 57	38, 40, 49, 52	46.75	44.75	+2.00
22	38, 40, 52, 57	39, 43, 48, 49	46.75	44.75	+2.00
23	38, 43, 49, 57	39, 40, 48, 52	46.75	44.75	+2.00
24	38, 39, 52, 57	40, 43, 48, 49	46.5	45	+1.50
25	38, 43, 48, 57	39, 40, 49, 52	46.5	45	+1.50
26	39, 40, 49, 57	38, 43, 48, 52	46.25	45.25	+1.00
27	38, 40, 49, 57	39, 43, 48, 52	46	45.5	+0.50
28	39, 40, 48, 57	38, 43, 49, 52	46	45.5	+0.50
29	38, 40, 48, 57	39, 43, 49, 52	45.75	45.75	0.00
30	38, 39, 49, 57	40, 43, 48, 52	45.75	45.75	0.00
31	38, 39, 48, 57	40, 43, 49, 52	45.5	46	−0.50
32	39, 40, 43, 57	38, 48, 49, 52	44.75	46.75	−2.00
33	38, 40, 43, 57	39, 48, 49, 52	44.5	47	−2.50
34	38, 39, 43, 57	40, 48, 49, 52	44.25	47.25	−3.00
35	38, 39, 40, 57	43, 48, 49, 52	43.5	48	−4.50

Mathematical recursion

Mathematical recursion, in a statistical context, is a process in which an initial probability value of a test statistic is calculated, then successive probability values are generated from the initial value by a recursive process. A recursive process is one in which items are defined in terms of items of similar kind. Using a recursive relation, a class of items can be constructed from one or a few initial values (a base) and a small number of relationships (rules). For example, given the base, $F_0 = 0$ and $F_1 = F_2 = 1$, the Fibonacci series $\{0, 1, 1, 2, 3, 5, 8, 13, 21, \ldots\}$ can be constructed by the recursive rule $F_n = F_{n-1} + F_{n-2}$ for $n > 2$. The initial value need not be an actual probability value, but can be a completely arbitrary positive value by which the resultant relative probability values are adjusted for the initializing value at the conclusion of the recursion process.

A recursion example

Consider a 2×2 contingency table using the notation in Table 3. Denote by a plus sign (+) the partial sum of all rows or all columns, depending on the position of the (+) in the subscript list. If the (+) is in the first subscript position, the sum is over all rows and if the (+) is in the second subscript position, the sum is over all columns. Thus, n_{i+} denotes the marginal frequency total of the ith row, $i = 1, \ldots r$, summed over all columns, n_{+j} denotes the marginal frequency total of the jth column, $j = 1, \ldots, c$, summed over

all rows, and $N = n_{11} + n_{12} + n_{21} + n_{22}$ denotes the Table 3 frequency total. The probability value corresponding to any set of cell frequencies in a 2×2 contingency table, $n_{11}, n_{12}, n_{21}, n_{22},$ is the hypergeometric point probability value given by

$$p = \binom{n_{+1}}{n_{11}}\binom{n_{+2}}{n_{12}}\binom{N}{n_{1}+}^{-1} = \frac{n_{1+}!\,n_{2+}!\,n_{+1}!\,n_{+2}!}{N!\,n_{11}!\,n_{12}!\,n_{21}!\,n_{22}!}. \quad (3)$$

Table 3: Conventional notation for a contingency table.

Category	Category		Total
	1	**2**	
1	n_{11}	n_{12}	n_{1+}
2	n_{21}	n_{22}	n_{2+}
Total	n_{+1}	n_{+2}	N

Since the exact probability value of a 2×2 contingency table with fixed marginal frequency totals and one degree of freedom is equivalent to the probability value of any one cell, determining the probability value of cell n_{11} is sufficient. If

$$p\{n_{11}+|n_{1+},n_{+1},N\} = p\{n_{11}+|n_{1+},n_{+1},N\} \times f(n_{11}), \quad (4)$$

Then, solving for $f(n_{11})$ produces

$$f(n_{11}) = \frac{p\{n_{11}+1|n_{1+},n_{+1},N\}}{p\{n_{11}|n_{1+},n_{+1},N\}} = \frac{n_{11}!\,n_{12}!\,n_{22}!}{(n_{11}+1)!(n_{12}-1)!(n_{22}-1)!} \quad (5)$$

and, after cancelling, yields

$$f(n_{11}) = \frac{n_{12}n_{21}}{(n_{11}+1)(n_{22}+1)}. \quad (6)$$

On 18 December 1934, R.A. Fisher presented an invited paper on statistical inference to the Royal Statistical Society, a paper that was subsequently published in the Journal of the Royal Statistical Society [10]. In this paper, Fisher described the logic of permutation statistical methods and provided an example based on 30 criminal same-sex twins from a study originally conducted by Dr. Johannes Lange, Chief Physician at the Munich-Schwabing Hospital in Schwabing, a northern suburb of Munich [11].

The Lange data analyzed by Fisher consisted of 13 pairs of monozygotic (identical) twins and 17 pairs of dizygotic (fraternal) twins. For each of the 30 pairs of twins, one twin was known to be a convict. The study considered whether the twin brother of the known convict was himself "convicted" or "not convicted," thus forming a 2×2 contingency table with 12 "convicted" and 18 "not convicted" twins cross-classified by the 13 "monozygotic" and 17 "dizygotic" twins. The 2×2 contingency table, as analyzed by Fisher, is presented in Table 4. Fisher determined the reference set of all possible arrangements of the four cell frequencies, given the observed marginal frequency totals. Since there is only one degree of freedom in Table 4, it was only necessary to determine the possible values for one cell. Consider cell n11 in the first row and first column of Table 4 with a frequency of 10. For the data in Table 4, there are $M = \min(n_{1.},n_{.1}) - \max(0,n_{11}-n_{22})+1$ possible, equally-likely arrangements of cell frequencies in the permutation reference set. Thus, for the

frequency data in Table 4.

Table 4: Convictions and non-convictions of same-sex twins.

Twin type	Convicted	Not convicted	Total
Monozygotic	10	3	13
Dizygotic	2	15	17
Total	12	18	30

$$M = \min(13,12) - \max(0,10-15)+1 = 12 - 0 + 1 = 13 \quad (7)$$

possible arrangements.

To begin a recursion procedure it is necessary to have an initial value; usually, a beginning probability value. However, it is not necessary to provide an actual probability value to initialize a recursion procedure. Any arbitrary positive value can serve as an initial value with a compensatory adjustment made at the conclusion of the recursion process. The recursion technique can be traced back at least to Lambert Adolphe Jacques Quetelet who used a recursion procedure to generate the binomial probability distribution with $p = 0.5$ and published the technique in a volume with the imposing title Letters Addressed to H.R.H. the Grand Duke of Saxe Coburg and Gotha on the Theory of Probabilities as Applied to the Moral and Political Sciences in 1846 [12]. To illustrate the use of an arbitrary origin in a recursion procedure, consider the data on monozygotic and dizygotic twins in Table 4 and set relative probability value $f\{n_{11}=0|13,12,30\}$ to a small arbitrarily-chosen value, say 1.00; thus, $f\{n_{11}=0|13,12,30\}=1.00.$ Then following Eq. (6), a recursion procedure produces

$$f\{n_{11}=1|13,12,30\} = 1.000 \times \frac{(13)(12)}{(1)(6)} = 26.000, \quad (8)$$

$$f\{n_{11}=2|13,12,30\} = 26.000 \times \frac{(12)(11)}{(2)(7)} = 245.1429, \quad (9)$$

$$f\{n_{11}=3|13,12,30\} = 245.1429 \times \frac{(11)(10)}{(3)(8)} = 1,123.5714, \quad (10)$$

$$f\{n_{11}=4|13,12,30\} = 1,123.5714 \times \frac{(10)(9)}{(4)(9)} = 2,808.9286, \quad (11)$$

$$f\{n_{11}=5|13,12,30\} = 2,8080.9286 \times \frac{(9)(8)}{(5)(10)} = 1,044.8571, \quad (12)$$

$$f\{n_{11}=6|13,12,30\} = 4,044.8571 \times \frac{(8)(7)}{(6)(11)} = 3,432.0000, \quad (13)$$

$$f\{n_{11}=7|13,12,30\} = 3,432.000 \times \frac{(7)(6)}{(7)(12)} = 1,716.000, \quad (14)$$

$$f\{n_{11}=8|13,12,30\} = 1,716.0000 \times \frac{(6)(5)}{(8)(13)} = 495.0000, \quad (15)$$

$$f\{n_{11}=9|13,12,30\} = 495.0000 \times \frac{(5)(4)}{(9)(14)} = 78.5714, \quad (16)$$

$$f\{n_{11}=10|13,12,30\} = 78.5714 \times \frac{(4)(3)}{(10)(15)} = 6.2857, \quad (17)$$

$$f\{n_{11}=11|13,12,30\} = 6.2857 \times \frac{(3)(2)}{(11)(16)} = 0.2143, \quad (18)$$

And

$$f\{n_{11}=12|13,12,30\} = 0.2143 \times \frac{(2)(1)}{(12)(17)} = 0.0021, \quad (19)$$

for a total of

$$T = \sum_{i=0}^{12} f\{n_{11}=i|13,12,30\} = 1.00 + 26.00 + \ldots + 0.0021 = 13,977.5735. \quad (20)$$

The desired exact hypergeometric point probability values are then obtained by dividing each relative probability value,

$f\{n_{11}|n_{1+},n_{+1},N\}$, by the recursively-obtained total, T.

For example,

$$p\{n_{11}=0|13,12,30\}=\frac{1.0000}{13,977.5735}=7.1543\times10^{-5}, \quad (21)$$

$$p\{n_{11}=1|13,12,30\}=\frac{26.0000}{13,977.5735}=1.8601\times10^{-3}, \quad (22)$$

$$p\{n_{11}=2|13,12,30\}=\frac{245.1429}{13,977.5735}=1.7538\times10^{-2}, \quad (23)$$

$$p\{n_{11}=3|13,12,30\}=\frac{1,123.5714}{13,977.5735}=8.0384\times10^{-2}, \quad (24)$$

$$p\{n_{11}=4|13,12,30\}=\frac{2,808.9286}{13,977.5735}=2.0096\times10^{-1}, \quad (25)$$

$$p\{n_{11}=5|13,12,30\}=\frac{4,044.8571}{13,977.5735}=2.8938\times10^{-1}, \quad (26)$$

$$p\{n_{11}=6|13,12,30\}=\frac{3,432.0000}{13,977.5735}=2.4554\times10^{-1}, \quad (27)$$

$$p\{n_{11}=7|13,12,30\}=\frac{1,716.0000}{13,977.5735}=1.277\times10^{-1}, \quad (28)$$

$$p\{n_{11}=8|13,12,30\}=\frac{495.0000}{13,977.5735}=3.5414\times10^{-2}, \quad (29)$$

$$p\{n_{11}=9|13,12,30\}=\frac{78.5714}{13,977.5735}=5.6212\times10^{-3}, \quad (30)$$

$$p\{n_{11}=10|13,12,30\}=\frac{6.2857}{13,977.5735}=4.4970\times10^{-4}, \quad (31)$$

$$p\{n_{11}=11|13,12,30\}=\frac{0.2143}{13,977.5735}=1.5331\times10^{-5}, \quad (32)$$

And,

$$p\{n_{11}=12|13,12,30\}=\frac{0.0021}{13,977.5735}=1.5030\times10^{-7}, \quad (33)$$

In this manner, the entire analysis is conducted utilizing an arbitrary initial value and a recursion procedure, thereby eliminating all factorial expressions. When the number of potential contingency tables given by is large, $M=\max(n_{11})-\min(n_{11})+1$ the computational savings can be substantial.

Computing the discrepancies from proportionality equal to or greater than the observed cell frequency configuration in Table 4, the one-tailed hypergeometric probability value for 10, 11, and 12 convicted monozygotic twins is

$$P=4.4970\times10^{-4}+1.5331\times10^{-5}+1.5030\times10^{-7}=4.6518\times10^{-4}. (34)$$

For the frequency data given in Table 4, a two-tailed hypergeometric probability value includes all hypergeometric point probability values equal to or less than the point.

Permutation Methods

Probability value of the observed contingency table, i.e., p{10|13, 12, 30} = 4.4970×10−4. In this case, the additional probability value associated with Table 4 is $p\{0|13,12,30\}=7.1534\times10^{-5}$. Thus, the two-tailed hypergeometric probability value is

$$P=4.4970\times10^{-4}+1.5331\times10^{-5}+1.5030\times10^{-7}+7.1543\times10^{-5}=5.3672\times10^{-4} (35)$$

Variable components of a test statistic

Under permutation, only the variable components of the test statistic need be calculated for each arrangement of the observed data. As this is often only a very small piece of the desired test statistic, calculations can often be reduced by several factors [13]. For example, in 1933 Thomas Eden and Frank Yates substantially reduced calculations in a randomized-block analysis of Yeoman II wheat shoots by recognizing that the block and total sums of squares would be constant for all of their 1,000 random samples and, consequently, the value of z for each sample would be uniquely defined by the treatment (between) sum of squares, i.e., the treatment sum of squares was sufficient for a permutation test of a randomized-block analysis of variance [5]. To illustrate an analysis using only the variable components of a test statistic, consider the expression for a conventional two-sample test t,

$$t=\frac{\overline{x}_1-\overline{x}_2}{\sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}, \quad (36)$$

where n_1 and n_2 denote the sample sizes, s_1^2 and s_2^2 denote the estimated population variances, $\overline{x}_1$ and $\overline{x}_2$ and denote the sample means for samples 1 and 2, respectively. In computing the permutation probability value of Student's two-sample t test, given the total of all response measurements

$$T=\sum_{i=1}^{n_1}x_{1i}+\sum_{i=1}^{n_2}x_{2i}, \quad (37)$$

Where, x_{1i} and x_{2i} denote the response measurements in samples 1 and 2, respectively, only the sum of the response measurements in the smaller of the two samples need be calculated, for each arrangement of the observed response measurements, i.e., $\sum_{i=1}^{n_1}x_{1i}$, where denotes a response measurement in sample 1 and $n_1\leq n_2$. Computing only the variable components of the test statistic thus eliminates a great deal of calculation for each random arrangement of the observed data, a time-saving technique utilized by E.J.G. Pitman in his 1937 permutation analysis of two independent samples. For a second example, consider Pearson's product-moment correlation coefficient between variables x and y given by

$$r_{xy}=\frac{\sum_{i=1}^{N}x_iy_i-\left(\sum_{i=1}^{N}x_i\sum_{i=1}^{N}y_i\right)\Big/N}{\sqrt{\left[\sum_{i=1}^{N}x_i^2-\left(\sum_{i=1}^{N}x_i\right)^2\Big/N\right]\left[\sum_{i=1}^{N}y_i^2-\left(\sum_{i=1}^{N}y_i\right)^2\Big/N\right]}}, \quad (38)$$

Where, N is the number of bivariate measurements. N and the summations

$$\sum_{i=1}^{N}x_i, \quad \sum_{i=1}^{N}x_i^2, \quad \sum_{i=1}^{N}y_i, \text{ and } \sum_{i=1}^{N}y_i^2 \quad (39)$$

are invariant under permutation. Thus, it is sufficient to calculate only $\sum_{i=1}^{N} x_i y_i$ for the observed data and all permutations of the observed data, eliminating a great deal of unnecessary calculation. In addition, it is only necessary to permute either variable x or variable y, leaving the other variable fixed.

For a third example, consider Kendall's τ_b coefficient of ordinal association. A number of measures of association for two ordinal-level variables are based on pairwise comparisons of differences between rank scores. The test statistic S, as defined by Maurice Kendall in 1938 and more extensively in 1948, plays an important role in a variety of statistical measures where Kendall's test statistic is often expressed as $S = C - D$, where C and D indicate the number of concordant pairs and discordant pairs, respectively [14, 15]. Consider two ordinal variables that have been cross-classified into a $r \times c$ contingency table, where r and c denote the number of rows and columns, respectively. Let n_{i+}, n_{+j}, and n_{ij} denote the row marginal frequency totals, column marginal frequency totals, and number of objects in the ij^{th} cell, respectively, for $i = 1, \ldots, r$ and $j = 1, \ldots, c$, and let N denote the total number of objects in the $r \times c$ contingency table, i.e.,

$$n_{i+} = \sum_{j=1}^{c} n_{ij}, \ n_{+j} = \sum_{j=1}^{r} n_{ij}, \text{ and } N = \sum_{i=1}^{r} \sum_{j=1}^{c} n_{ij}. \quad (40)$$

Kendall's τ_b coefficient of ordinal association between variables x and y given by

$$\tau_b = \frac{S}{\sqrt{(C+D+T_x)(C+D+T_y)}}, \quad (41)$$

Where, for an $r \times c$ contingency table C denotes the number of pairs of objects that are ranked in one order on both variable x and y given by

$$C = \sum_{i=1}^{r-1} \sum_{j=1}^{c-1} n_{ij} \left(\sum_{k=i+1}^{r} \sum_{l=j+1}^{c} n_{kl} \right), \quad (42)$$

D denotes the number of pairs of objects that are ranked in one order on variable x and the reverse order on variable y given by

$$D = \sum_{i=1}^{r-1} \sum_{j=1}^{c-1} n_{ij} \left(\sum_{k=i+1}^{r} \sum_{l=j+1}^{c} n_{kl} \right), \quad (43)$$

T_x denotes the number of pairs tied on variable x but not tied on variable y given by

$$T_x = \sum_{i=1}^{r} \sum_{j=1}^{c-1} n_{ij} \left(\sum_{k=j+1}^{c} n_{ik} \right), \quad (44)$$

And T_y denotes the number of pairs tied on variable y but not tied on variable x given by

$$T_y = \sum_{j=1}^{c} \sum_{i=1}^{r-1} n_{ij} \left(\sum_{k=j+1}^{r} n_{kj} \right). \quad (45)$$

To illustrate the calculation of Kendall's τ_b test statistic, consider the frequency data given in Table 5, where $N = 41$ bivariate observations have been cross-classified into a 3×3 ordered contingency table. For the frequency data given in Table 5, the number of concordant pairs following Eq. (42) is

Table 5: Example rank-score data for bivariate observations cross-classified on ordinal variables and into a contingency table.

x	y			Total
	1	**2**	**3**	
1	7	6	3	16
2	5	2	7	14
3	3	2	6	11
Total	15	10	16	41

C=(7)(2 +7+2+6)+(6)(7+6)+(5)(2+ 6)+(2)(6)=249, (46)

The number of discordant pairs following Eq. (43) is

D=(3)(5+2+3+2)+(6)(5+3)+(7)(3+2)+(2)(3)=125, (47)

The number of pairs tied on variable x following Eq. (44) is

T_x=(7)(6+3)+(6)(3)+(5)(2+7)+(2)(7)+(3)(2+6)+(2)(6)=176, (48)

The number of pairs tied on variable y following Eq. (45) is

T_y=(7)(5+3)+(5)(3)+(6)(2+2)+(2)(2)+(3)(7+6)+(7)(6)=180, (49)

S =CD=249 -125=+124, and the observed value of Kendall's τ_b test statistic is

$$\tau_b = \frac{S}{\sqrt{(C+D+T_x)(C+D+T_y)}} = \frac{+124}{\sqrt{(249+125+176)(249+125+180)}} = +0.2246.$$

It is not generally recognized that $C+D+T_x$ and $C+D+T_y$ can be obtained from the observed marginal frequency totals, which are invariant under permutation. Thus,

$$C+D+T_x = \frac{1}{2} \left(N^2 - \sum_{j=1}^{c} n_{+j}^2 \right) \quad (51)$$

And

$$C+D+T_y = \frac{1}{2} \left(N^2 - \sum_{j=1}^{c} n_{i+}^2 \right). \quad (52)$$

Since, the marginal frequency distributions are fixed under permutation, the exact probability value of Kendall's τ_b is based entirely on the permutation distribution of S. Not only does this hold for Kendall's τ_b, but it also holds for other test statistics with in the numerator, including Kendall's τ_a, Stuart's τ_c, and Somers' d_{yx} and d_{xy} measures of ordinal association.

For an exact permutation analysis of an $r \times c$ contingency table, it is necessary to calculate the selected measure of ordinal association for the observed cell frequencies and exhaustively enumerate all M possible, equally-likely arrangements of the N objects in the rc cells, given the observed marginal frequency distributions. For each arrangement in the reference set of all permutations of cell frequencies a measure of ordinal association, say S, and the exact hypergeometric point probability value under the null hypothesis, $p\left(n_{ij} \middle| n_{i+}, n_{+j}, N\right)$, are calculated,

Where,

$$p\left(n_{ij} \middle| n_{i+}, n_{+j}, N\right) = \frac{\left(\prod_{i=1}^{r} n_{i+}!\right)\left(\prod_{j=1}^{c} n_{+j}!\right)}{N \prod_{i=1}^{r} \prod_{j=1}^{c} n_{ij}!}, \quad (53)$$

n_{ij} is an observed cell frequency for $i = 1, \dots r$ and $j = 1, \dots, c, n_{i+}$ is the ith of row marginal frequency totals summed over all columns, n_{+j} is the jth of column marginal frequency totals summed over all rows, and N is the total of all n_{ij} values for $i = 1, \dots r$ and $j = 1, \dots, c$. If S_o denotes the value of the observed test statistic, the exact one-sided upper- and lower-tail probability (P) values of S_o are the sums of the $p\left(n_{ij} \middle| n_{i+}, n_{+j}, N\right)$ values associated with the S values computed on all possible, equally-likely arrangements of cell frequencies that are equal to or greater than S_o when S_o is positive and equal to or less than So when So is negative, respectively. Thus, the exact hypergeometric probability value of S_o when S is positive is given by

$$P = \sum_{k=1}^{M} \Psi\left(S_k\right) p\left(n_{ij} \middle| n_{i+}, n_{+j}, N\right), \quad (54)$$

Where,

$$\Psi\left(S_k\right) = \begin{cases} 1 & \text{if } S_k \geq S_o, \\ 0 & \text{otherwise}, \end{cases} \quad (55)$$

and the exact hypergeometric probability value of S_o when S is negative is given by

$$P = \sum_{k=1}^{M} \Psi\left(S_k\right) p\left(n_{ij} \middle| n_{i+}, n_{+j}, N\right), \quad (56)$$

Where,

$$\Psi\left(S_k\right) = \begin{cases} 1 & \text{if } S_k \leq S_o, \\ 0 & \text{otherwise.} \end{cases} \quad (57)$$

For the frequency data given in Table 5, there are only $M = 5,225$ possible, equally-likely arrangements in the reference set of all permutations of cell frequencies given the observed row and column marginal frequency distributions, $\{16,14,11\}$ and $\{15,10,16\}$, respectively, making an exact permutation analysis possible. The exact probability value of Kendall's S under the null hypothesis is the sum of those hypergeometric

point probability values associated with values of $S = +124$ or greater. Based on the hypergeometric probability distribution, the exact upper-tail probability value is $P = 0.0555$.

Limited number of discrete values

Simple linear correlation with one predictor variable requires the enumeration of $N!$ possible arrangements of the observed data when N determining an exact probability value. When is large, the number of possible arrangements can be excessive. For example, with only $N = 15$ paired values there are

$$M = 15! = 1,307,674,368,000 \quad (58)$$

arrangements to be enumerated and analyzed. In such cases, Monte Carlo resampling permutation methods can be utilized, in which a random sample of size L is drawn from the reference set of all M possible arrangements. The probability value is simply the proportion of sampled correlation coefficients equal to or more extreme than the observed correlation coefficient value. While the resulting probability value is approximate, when L is large the probability value can be highly accurate. In general, a sample of $L = 1,000,000$ random arrangements is sufficient to ensure three decimal place of accuracy. However, $L = 100,000,000$ random arrangements is required to ensure four decimal places of accuracy.

Under special circumstances, there is an alternative for determining an exact probability value. Occasionally, correlation data is gathered in which one of the variables has only a limited number of unique values. The limiting case occurs when, for example, y variable is a continuous interval-level variable and variable x is a simple dichotomy, such as in point-biserial correlation. However, there are other cases in which only a few distinct values appear in variable , e.g., three or four values. An exact analysis can then be conducted by treating the distinct values of variable x as groups and correlating the continuous variable with group membership, with considerable increase in calculation efficiency.

To illustrate the analysis technique, consider $N = 12$ paired x and y values where variable x is a simple dichotomy. The number of objects in group 1 are indicated by $n_1 = 6$ and the number of objects in group 2 are indicated by $n_2 = 6$. There are

$$M = N! = 12! = 479,001,600 \quad (59)$$

possible arrangements of the observed data to be enumerated and analyzed. However, treating the data as consisting of two groups with $n_1 = n_2 = 6$, there are only

$$M = \frac{N!}{n_1! n_2!} = \frac{12!}{6!6!} = 924 \quad (60)$$

arrangements to be examined. If, for example, the groups are not of equal size and $N = 12$, $n_1 = 7$ and $n_2 = 65$, then there are only

$$m = \frac{N!}{n_1! n_2!} = \frac{12!}{7!5!} = 792 \quad (61)$$

arrangements to be examined. If $N = 12$ and there are three distinct values of variable x where $n_1 = n_2 = n_3 = 4$, then there are

$$m = \frac{N!}{n_1!n_2!n_3!} = \frac{12!}{4!4!4!} = 34,650 \qquad (62)$$

arrangements to be examined. As with a simple dichotomy, as the group sizes diverge, the number of possible arrangements to be examined is further reduced. For example, if $N = 12$, $n_1 = 3, n_2 = 4,$ and $n_3 = 5,$ then there are only

$$m = \frac{N!}{n_1!n_2!n_3!} = \frac{12!}{3!4!5!} = 27,720 \qquad (63)$$

arrangements to be examined.

Two examples illustrate the computational efficiency of analyzing group structures. First, consider the correlation data listed in Table 6. The data in Table 6 consist of $N = 12$ subjects divided into two groups with seven Republicans (R) in group 1 and five Democrats (D) in group 2. The raw data are listed on the left side of Table 6 and the dummy-coded (0, 1) values are listed on the right side. A dummy code of 1 indicates that the subject is in group 1, i.e, is a Republican, and a dummy code of 0 indicates that the subject is not in group 1, i.e., is a Democrat.

Table 6: Twelve subjects with paired and values listed on the left side and dummy-coded, rearranged values on the right side.

Subject	x	y	Subject	x	y
1	R	1	1	0	1
2	D	2	2	0	3
3	R	3	3	0	4
4	R	4	4	0	5
5	R	5	5	0	6
6	R	6	6	0	8
7	D	7	7	0	11
8	R	8	8	1	2
9	D	9	9	1	7
10	D	10	10	1	9
11	D	11	11	1	10
12	D	12	12	1	12

For the paired data listed in Table 1, $N = 12$, $n_1 = 7, n_2 = 5,$ and the product-moment correlation between variables x and y is $r_{xy} = +0.3672$. Treating variable x as two groups, there are only

$$m = \frac{N!}{n_1!n_2!} = \frac{12!}{7!5!} = 792 \qquad (64)$$

arrangements to be examined. There are $212 r_{xy}$ values equal to or more extreme than the observed value of $r_{xy} = +0.3672$. If r_o denotes the observed value of the exact two-sided probability value of $r_o = +0.3672$ is

$$P = \frac{\text{number of } |r_{xy}| \text{ values } \geq |r_0|}{m} = \frac{212}{792} = 0.2677. \qquad (65)$$

In contrast, for an exact analysis of all
$$M = N! = 12! = 479,001,600 \qquad (66)$$

arrangements of the observed data, there are $128,220,580 r_{xy}$ values equal to or more extreme than the observed value of $r_{xy} = +0.3672$. If denotes the observed value of r_{xy}, the exact two-sided probability value of $r_o = +0.3672$ is

$$P = \frac{\text{number of } |r_{xy}| \text{ values } \geq |r_0|}{m} = \frac{128,220,2585}{479,001,600} = 0.2677. \qquad (67)$$

For a second example, consider the correlation data listed in Table 7. The data in Table 7 consist of $N = 12$ subjects divided into three groups with four Republicans (R) in group 1, four Democrats (D) in group 2, and four independents (I) in group 3. The raw data is listed on the left side of Table 7 and the dummy-coded (0, 1) values are listed on the right side. A dummy code of (1, 0) indicates that the subject is in group 1, i.e., is a Republican, a dummy code of (0, 1) indicates that the subject is not in group 1, but is in group 2, i.e., is a Democrat, and a dummy code of (0, 0) indicates that the subject is not in groups 1 or 2, but is in group 3, i.e., is an Independent.

Table 7: Twelve subjects with x_1, x_2, y and values listed on the left side and dummy-coded, rearranged values on the right side.

Subject	x	y	Subject	x_1	x_2	y
1	I	1	1	1	0	4
2	I	2	2	1	0	9
3	D	3	3	1	0	10
4	R	4	4	1	0	11
5	I	5	5	0	1	3
6	I	6	6	0	1	7
7	D	7	7	0	1	8
8	D	8	8	0	1	12
9	R	9	9	0	0	1
10	R	10	10	0	0	2
11	R	11	11	0	0	5
12	D	12	12	0	0	6

For the paired data listed in Table 7, $N = 12$, $n_1 = n_2 = n_3 = n_4,$ and the product-moment correlation between the linear combination of variables x_1 and x_2 with variable y is $R_{y.x_1,x_2} = +0.3916$. Treating variable x as three groups, there are

$$m = \frac{N!}{n_1!n_2!n_3!} = \frac{12!}{4!4!4!} = 34,650 \qquad (68)$$

arrangements to be examined. There are $3,948 R_{y.x_1,x_2}$, values equal to or more extreme than the observed value of $R_{y.x_1,x_2} = +0.3916$. If R_o denotes the observed value of $R_{y.x_1,x_2}$, the exact two-sided probability value of $R_{y.x_1,x_2,} = +0.3916$ is

$$P = \frac{\text{number of } \mathrm{R}_{y.x_1,x_2} \text{ values } \geq r_0}{m} = \frac{3,948}{34,650} = 0.1139. \quad (69)$$

In contrast, for an exact analysis of all

$$M = N! = 12! = 479,001,600 \quad (70)$$

arrangements of the observed data, there are $54,576,484 R_{y.x_1,x_2}$ values equal to or more extreme than the observed value of $R_{y.x_1,x_2} = 0.3916$. R_o denotes the observed value of $R_{y.x_1,x_2}$, the exact two-sided probability value of $R_{y.x_1,x_2} = 0.3916$ is

$$P = \frac{\text{number of } \mathrm{R}_{y.x_1,x_2} \text{ values } \geq r_0}{m} = \frac{54,576,484}{479,001,600} = 0.1139. \quad (71)$$

In this manner, when there is a limited number of distinct values in one or more variables, the number of possible arrangements of the observed data can be reduced substantially and an exact analysis conducted where an examination of all $N!$ arrangements would be prohibitive.

Fixed measurements

In matched-pairs and other block designs where the same, or matched, subjects are measured on two or more occasions, the number of possible arrangements of the observed data can become quite large. Holding one set of measurements constant, relative to the other set(s), can greatly reduce the number of arrangements to be analyzed. In 1939 Maurice Kendall and Bernard Babington Smith published an article in The Annals of Mathematical Statistics on "The problem of m rankings" in which they developed the well-known coefficient of concordance [16]. Let N and m denote the number of rank scores and the number of judges, respectively, then Kendall and Babington Smith defined the coefficient of concordance as

$$W = \frac{12S}{m^2 \left(N^3 - N \right)}, \quad (72)$$

Where, S is the observed sum of squares of the deviations of sums of ranks from the mean value $m(N++)/2$. In a form more conducive to calculation, W can be defined as

$$W = \frac{12 \sum_{i=1}^{N} R_i^2 - 3m^2 N (N+1)}{m^2 N \left(N^2 - 1 \right)}, \quad (73)$$

Where, R_i for $i = 1, \ldots, N$ is the sum of the rank scores for the ith of N objects and there are no tied rank scores. Since $m^2 \left(N^3 - N \right)$ in the denominator of Eq. (72) is invariant over all permutations of the observed data, Kendall and Babington Smith showed that in order to test whether an observed value of W is statistically significant it is only necessary to consider the distribution of S by permuting the N ranks in all possible, equally-likely ways. There are

$$M = \left(N! \right)^m \quad (74)$$

possible arrangements of the observed data to be analyzed. However, letting one of the rankings be fixed, relative to the other $m-1$ rankings, there are only

$$M = \left(N! \right)^{m-1} \quad (75)$$

arrangements to be considered.

Example

To illustrate Kendall and Babington Smith's coefficient of concordance, consider the rank data listed in Table 8 with $N = 8$ objects and $m = 3$ sets of rankings. For the rank scores listed in Table 8, the sum of the squared rank scores is

Table 8: Example data for Kendall and Babington Smith's coefficient of concordance with $m = 3$ sets of rankings and $N = 8$ objects.

Object	Ranking			R_i
	1	2	3	
1	4	2	5	11
2	6	6	3	15
3	2	1	1	4
4	1	3	2	6
5	3	5	8	16
6	5	4	4	13
7	7	8	7	22
8	8	7	6	21

$$\sum_{i=1}^{N} R_i^2 = 11^2 + 15^2 + 4^2 + 6^2 + 16^2 + 132 + 22^2 + 21^2 = 1,748, \quad (76)$$

the observed value of Kendall and Babington Smith's coefficient of concordance is

$$W = \frac{12 \sum_{i=1}^{N} R_i^2 - 3m^2 N(N+1)}{m^2 N \left(N^2 - 1 \right)} = \frac{12(1,748) - (3)(3^2)(8)(8+1)^2}{3^2 (8)(8^2 - 1)} = 0.7672, \quad (77)$$

For the $m = 3$ sets of rank scores listed in Table 7 there are

$$M = \left(N! \right)^m = \left(8! \right)^3 = 65,548,320,768,000 \quad (78)$$

possible arrangements of the observed data- too many for an exact permutation analysis. However, holding one set of rank scores constant relative to the other two sets of rank scores reduces the number of possible arrangements to

$$M = \left(N! \right)^m = \left(8! \right)^3 = 1,625,702,400 \quad (79)$$

While even this number of possible arrangements makes an exact permutation analysis impractical, it is not impossible. If the reference set of all M arrangements of the observed rank scores listed in Table 8 occur with equal chance, the exact probability of $W = 0.7672$ under the null hypothesis is

$$P\left(W \geq W_o | H_0 \right) = \frac{\text{number of } W \text{ values } \geq W_0}{M} = \frac{5,125,594}{1,625,702,400} = 0.3153 \times 10^{-2}, \quad (80)$$

Where, W_o denotes the observed value of W.

Partition theory

Goodness-of-fit tests are common in contemporary research. Among the more popular goodness-of-fit tests are Pearson's

chi-squared test (χ^2) [17], Wilks' likelihood-ratio test (G^2) [18,19], the Freeman-Tukey test (T^2) [20], and the Cressie–Read test $(I(\lambda))$ [4]. In this section goodness-of-fit tests for unordered equiprobable categories are described and compared. Included in this section are Fisher's exact test, exact chi-squared, exact likelihood-ratio, exact Freeman-Tukey, and exact Cressie-Read goodness-of-fit tests for k disjoint, unordered categories with equal probabilities under the null hypothesis. Exact tests are free from any asymptotic assumptions; consequently, they are ideal for sparse tables where expected values may be small. Consider the random assignment of N objects to k unordered, mutually-exclusive, exhaustive, equiprobable categories, i.e., the probability for each of the k categories is $p_i = 1/k$ for $i = 1, \ldots, k$ under the null hypothesis. Then, the probability that O_i objects occur in the ith of k categories is the multinomial probability given by

$$P\left(O_i \mid p_i, N\right) = P\left(O_1, \ldots, O_k \mid p_1, \ldots, p_k, N\right) = \left(N \middle/ \prod_{i=1}^{k} O_i! \right) \prod_{i=1}^{k} p_i^{O_i}, \quad (81)$$

Fisher's exact goodness-of-fit test is the sum of all distinct $P\left(O_i \mid N, p_i\right)$ values that are equal to or less than the observed value of $P\left(O_i \mid N, p_i\right)$ associated with a set of observations, $O_1, \ldots, O_k$ [21]. The Pearson chi-squared goodness-of-fit test statistic for N objects in k disjoint, unordered categories is given by

$$\chi^2 = \sum_{i=1}^{k} \frac{\left(O_i - E_i\right)^2}{E_i} \quad (83)$$

and Wilks' likelihood-ratio test statistic is given by

$$G^2 = 2\sum_{i=1}^{k} O_i In\left(\frac{O_i}{E_i}\right), \quad (84)$$

Where, the expected frequency of the ith category under the null hypothesis of equal category probabilities is given by

$$E_i = \frac{N}{k} \text{ for } i = 1, \ldots, k. \quad (85)$$

Two other tests that have received attention are the Freeman-Tukey [9] goodness-of-fit test given by

$$T^2 = \sum_{i=1}^{k} \left[\sqrt{O_i} + \sqrt{O_i + 1} - \sqrt{4N/k + 1} \right]^2 \quad (86)$$

and the Cressie-Read [4] goodness-of-fit test given by

$$I\left(\lambda\right) = \frac{2}{\lambda\left(\lambda + 1\right)} \sum_{i=1}^{k} O_i \left[\left(\frac{kO_i}{N}\right)^{\lambda} - 1 \right]. \quad (87)$$

Cressie and Read showed that $I(\lambda)$ with λ set to 2/3 was optimal both in terms of attained significance level and small sample properties. In general, for an exact goodness-of-fit test with N objects in k categories there are

$$M = \binom{N + k - 1}{k - 1} \quad (88)$$

distinct ordered arrangements to be examined. Under the null hypothesis that the probabilities of all k categories are equal, a vastly reduced number of distinct partitions of the data can be considered using a further condensation of the M ordered configurations. The condensation is based on a 1748 result by Leonhard Euler that provides a generating function for the number of decompositions of N integer summands without regard to order using the recurrence relation

$$p\left(N\right) = \sum_{j=1}^{} \left(-1\right)^{j-1} p\left[N - \frac{3j^2 \pm j}{2} \right], \quad (89)$$

Where, $p(0) = 1$ and   is a positive integer satisfying $2 \le 3j^2 \pm j \le 2N$. Note that if $N = 1$, then $j = 1$, with only the minus (−)sign allowed; if $2 \le N \le 4$, then $j = 1$, with both the plus (+) and minus (−) signs allowed; if $5 \le N \le 6$, then $j = 1$ with both the plus (+) and minus (−) signs allowed and $j = 2$ with only the minus (−) sign allowed; and so forth. G.H. Hardy and S. Ramanujan [22] provided the asymptotic formula for $p(N)$ given by

$$p\left(N\right) \sim \frac{1}{4N\sqrt{3}} \exp\left(\pi\sqrt{2N/3}\right) \quad (90)$$

as $N \to \infty$.

Given   disjoint, unordered categories and observed categorical frequencies $O_1, \ldots, O_k$, an algorithmic procedure generates all $p(N)$ partitions, computes the exact probability for each partition, calculates the observed $\chi^2, G^2, T^2,$ and $I(\lambda)$ test statistic values, and calculates the number of ways that each partition can occur, i.e., the partition weights [23]. The partition weights are multinomial and are given by

$$W = \frac{k!}{\prod_{i=1}^{m} f_i!}, \quad (91)$$

Where, f_i is the frequency for each of m distinct integers comprising a partition. For example, if the observed partition $\{3\,2\,2\,1\,0\,0\}$ is where $N = 8$ objects, $k = 6$ categories, $m = 4$ and distinct integers $(3, 2, 1, 0)$, then $f_1 = 1$, $f_2 = 2$, $f_3 = 1$, $f_4 = 2$, and

$$W = \frac{6!}{(1)(2)(1)(2)} = \frac{720}{4} = 180. \quad (92)$$

If $k < N$, the number of distinct partitions is reduced to eliminate those partitions where the number of partition values exceeds k. For example, if $k = 3$ and $N = 5$, then the two partitions $\{2\,1\,1\}$ and $\{1\,1\,1\,1\,1\}$ cannot be considered as the respective number of partitions, four and five, both exceed $k = 3$. The sum of the values of W for the included distinct partitions is equal to M. The exact probability values for the Fisher exact, Pearson χ^2, Wilks G^2, Freeman-Tukey T^2, and Cressie-Read $I(\lambda)$ goodness-of-fit tests are obtained by comparing observed values to partition probability values. In the case of Fisher's exact test, partition probability values equal to or less than the observed partition probability value are weighted and summed.

For the exact χ^2, G^2, T^2, and $I(\lambda)$ goodness-of-fit tests, partition probability values associated with partition test statistics equal to or greater than the observed test statistic values are weighted and summed.

Example

Consider that a patient is asked to check any of $k = 50$ symptoms experienced in the past six months, resulting in $N = 10$ selections for categorical frequencies of $O_1 = 4, O_2 = 3, O_3 = 2, O_4 = 1$, and $O_5 = \ldots = O_{50} = 0$. The null hypothesis specifies that the k expected category probabilities are equally likely, i.e., $p_i = 1/k = 1/50 = 0.02$ for $i = 1, \ldots, 50$. Fisher's $i = 1, \ldots, 50$. exact probability value is 0.1795×10^{-5}. Under the null hypothesis of no difference between the observed and expected cell frequencies, the observed Pearson uncorrected chi-squared test statistic is $\chi^2 = 140.00$ with an exact probability value of $P = 0.3788 \times 10^{-4}$, the observed Wilks likelihood-ratio test statistic is $G^2 = 52.64$ with an exact probability value of $P = 0.1795 \times 10^{-5}$, the observed Freeman–Tukey test statistic is $T^2 = 23.87$ with an exact probability value of $P = 0.1795 \times 10^{-5}$, and the observed Cressie-Read test statistic is $I(2/3) = 89.87$ with an exact probability value of $P = 0.8356 \times 10^{-5}$. Table 9 lists the exact probability values for the Fisher exact, Pearson χ^2, Wilks G^2, Freeman-Tukey T^2, and Cressie-Read $I(2/3)$ test statistics and illustrates the analysis of very sparse data with $N = 10$ observations in $k = 50$ unordered equiprobable categories[24-26].

Table 9: Test statistics and exact probability values for the Fisher exact, Pearson χ^2, Wilks G^2, Freeman–Tukey T^2, and Cressie–Read $I(2/3)$ tests for $N = 45$, $k = 20$, and $p_i = 0.05$ for $i = 1, \ldots k$.

Test	Statistic	Probability
Fisher exact test	–	0.1795×10^{-5}
Pearson χ^2	140.00	0.3788×10^{-4}
Wilks G^2	52.64	0.1795×10^{-5}
Freeman–Tukey T^2	23.87	0.1795×10^{-5}
Cressie–Read $I(2/3)$	89.87	0.8356×10^{-5}

For the example analyses in Table 9 there are only $p(10) = 42$ partitions of the observed data to be examined. On the other hand, for a conventional analysis there are

$$M = \binom{N+k-1}{k-1} = \binom{10+50-1}{50-1} = 62,928,356,305 \qquad (93)$$

arrangements to be examined. The use of partitions based on $p(N)$ with equal category probabilities is highly efficient when compared with the calculation of test statistic values based on all M possible configurations. Table 10 compares the $p(N)$ partitions with the M possible configurations for $1 \le N = k \le 20$. For example, with $N = 15$ observations in $k = 15$ categories, there are only $p(15) = 176$ partitions, but

$$M = \binom{N+k-1}{k-1} = \binom{15+15-1}{15-1} = \binom{29}{14} = \frac{29!}{14!15!} = 77,558,760 \qquad (94)$$

Table 10: Comparison of $p(N)$ partitions and M configurations with $1 \le N = k \le 20$.

N	$p(N)$	M
1	1	1
2	2	3
3	3	10
4	5	35
5	7	126
6	11	462
7	15	1,716
8	22	6,435
9	30	24,310
10	42	92,378
11	56	352,716
12	77	1,352,078
13	101	5,200,300
14	135	20,058,300
15	176	77,558,760
16	231	300,540,195
17	297	1,166,803,110
18	385	4,537,567,650
19	490	17,672,631,900
20	627	68,923,264,410

Monte carlo resampling

When exact permutation procedures become unmanageable due to the large number of possible arrangements of the observed data, a random sample of the M possible arrangements can be substituted, providing approximate, but highly accurate, probability values. Resampling permutation tests generate and examine a Monte Carlo random sample of all possible, equally-likely arrangements of the observed response measurements. For each randomly selected arrangement of the observed data, the desired test statistic is calculated. The probability of obtaining the observed value of the test statistic, or one more extreme, is the proportion of the randomly selected test statistics with values equal to or more extreme than the value of the observed test statistic. With a sufficient number of random samples, a probability value can be computed to any reasonable accuracy. While exact permutation methods are restricted to relatively small samples, given the large number of possible arrangements of the observed data, Monte Carlo resampling permutation tests are not limited by the sizes of the samples.

Monte Carlo resampling permutation tests have been shown to provide good approximations to exact probability values as a function of the number of random arrangements considered. The number of random arrangements suggested in books and articles is varied and likely dated due to previous limitations of computer speed and memory. Some authors have proposed as few as 100 random arrangements to as many as

10,000. For many years, 10,000 random arrangements was the most common number cited and the default value in many commercial resampling programs. More recently, 1,000,000 random arrangements appears to be the number most in use [13]. To illustrate the number of random arrangements required to yield a predetermined number of decimal places of accuracy, consider the data listed in Table 11. The data listed in Table 11 are adapted from Berry, Mielke, and Mielke and represent soil lead (Pb) quantities from two school districts in metropolitan New Orleans. There are

Table 11: Ordered soil Pb data in mg/kg from two school attendance districts in metropolitan New Orleans.

n	District	
	1	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	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	21,679.0

$$M = \frac{(n_1 + n_2)!}{n_1! \, n_2!} = \frac{(20+20)!}{20! \, 20!} = 137,846,528,820 \quad (95)$$

possible arrangements of the soil lead data listed in Table 11 to be considered. Table 12 summarizes the results for eight different resamplings (L) of the data listed in Table 11 and the associated two-sided probability values. Each of the probability values was generated using a common seed and the same pseudorandom number generator. The last row of Table 12 contains the exact probability value based on all $M = 137,846,528,820$ possible arrangements of the soil lead data listed in Table 11. Given the results of the Monte Carlo resampling probability analyses listed in Table 12, $L = 1,000,000$ appears adequate when three decimal places of accuracy are required. Little improvement is achieved with increasing numbers of permutations. $L = 1,000,000$ provides

a combination of accuracy and practicality. Modern desktop computers can generate 1,000,000 random arrangements in just a few seconds. For a second example analysis of a Monte Carlo resampling analysis consider the frequency data given in Table 13 where $N = 72$ bivariate observations have been cross-classified into a 3×5 ordered contingency table.

Table 12: Comparison of eight resampled probability values with the exact probability value given in the last row.

Resampling (L)	Probability (P)
1,000	0.02
10,000	0.011
100,000	0.01556
1,000,000	0.014946
10,000,000	0.0149302
100,000,000	0.0148851
1,000,000,000	0.014917218
137,846,528,820	0.014918212

To calculate an exact probability value for a contingency table it is necessary to exhaustively enumerate all possible arrangements of cell frequencies, given the observed marginal distributions, calculate the specified test statistic on each arrangement, compute the hypergeometric point probability value for each arrangement of the observed data, and sum the hypergeometric point probability values associated with those test statistic values that are equal to or more extreme than the test statistic value calculated on the observed contingency table. A hypergeometric probability value, $p\left(n_{ij} \mid n_{i+}, n_{+j}, N\right)$, is given by

$$p\left(n_{ij} \mid n_{i+}, n_{+j}, N\right) = \frac{\left(\prod_{i=1}^{r} n_{i+}!\right)\left(\prod_{j=1}^{c} n_{+j}!\right)}{N \prod_{i=1}^{r} \prod_{j=1}^{c} n_{ij}!}, \quad (96)$$

Where, n_{ij} is an observed cell frequency for $i = 1,\dots r$ and $j = 1,\dots c, n_{i+}$ is the ith of row marginal frequency totals summed over all columns, n_{+j} is the jth of column marginal frequency totals summed over all rows, and N is the total of all n_{ij} values for $i = 1,\dots r$ and $j = 1,\dots c$. On the other hand, Monte Carlo resampling only requires counting the number of test statistic values that are equal to or greater than the observed test statistic value-no hypergeometric probability values are required. The Monte Carlo resampling probability value is simply the proportion of test statistic values computed on the randomly selected arrangements that are equal to or greater than the observed test statistic value. For the frequency data given in Table 13, Kendall's measure of ordinal association as given in Eq. (41) is $\tau_b = +0.1664$. For the data given in Table 13, there are $M = 70,148,145$ possible, equally-likely arrangements in the reference set of all permutations of cell frequencies given the observed row and column marginal frequency distributions, $\{23, 23, 26\}$ and $\{15,11,14,17,15\}$, respectively. Therefore, an exact test

may not be practical and a Monte Carlo resampling probability analysis based on L random arrangements of cell frequencies is utilized. The resampling probability value of τ_b under the null hypothesis isthe number of τ_b values equal to or greater than $\tau_b = +0.1664$; in this case the Monte Carlo resampling approximate upper-tail probability value based on $L = 1,000,000$ is

$$P\left(\tau_b \geq \tau_O\right) = \frac{\text{number of } \tau_b \text{ values } \geq \tau_O}{L} = \frac{48,600}{1,000,000} = 0.0486, \quad (97)$$

Table 13: Example data for $N = 72$ bivariate observations cross-classified on ordinal variables x and y into a 3×5 contingency table.

x	y					Total
	1	**2**	**3**	**4**	**5**	
1	8	4	3	5	3	23
2	3	2	8	5	5	23
3	4	5	3	7	7	26
Total	15	11	14	17	15	72

Where, τ_O denotes the observed value of τ_b. While an exact permutation analysis is not practical for the frequency data given in Table 1, it is not impossible. The exact probability value of Kendall's τ_b under the null hypothesis is the sum of those hypergeometric point probability values associated with values of $\tau_b = +0.1664$ or greater. Based on the hypergeometric probability distribution, the exact upper-tail probability value conditioned on the reference set of all $M = 70,148,145$ arrangements of cell frequencies is

$$P\left(\tau_b \geq \tau_O\right) = \frac{\text{number of } \tau_b \text{ values } \geq \tau_O}{L} = \frac{3,423,229}{70,148,145} = 0.0488, \quad (98)$$

A Monte Carlo resampling analysis not only examines a small fraction of the M possible arrangements of the observed data but, as in this example analysis, eliminates the calculation of 70,148,145 hypergeometric point probability values for a substantial increase in calculation efficiency.

Discussion

Eight improvements to permutation statistical methods are described to improve calculation efficiency. First, high-speed computers can quickly generate a large number of arrangements of the observed data, be these all possible arrangements or a large number of random arrangements. Second, substitution of all possible combinations for permutations yields the same probability values with a great reduction in computing time. Third, mathematical recursion greatly simplifies difficult calculations. Fourth, analysis of only the variable components of a specified test statistic eliminates many non-essential calculations. Fifth, when one variable contains only a small number of different values, substantial savings can be achieved by treating the different values as groups with no loss of precision. Sixth, For certain classes of problems, e.g., block designs, fixing one of the variables relative to the other variables can reduce the number of arrangements that need to be examined. Seventh, the use of partition theory in goodness-of-fit tests greatly reduces the number of required arrangements of the observed data. Eighth, Monte Carlo resampling methods can provide highly-accurate approximate probability values for samples of any size with substantial reduction in calculation. The first seven are methods to improve calculation of exact probability values, but may also be used with Monte Carlo resampling values. While any one of the eight improvements may be used independently of the others, when used in combination with one or more of the others, vastly improved computation times can be achieved.

References

1. Fisher RA (1925) Statistical Methods for Research Workers. In: Oliver B, Edinburgh, Biological Monographs and Manuals, Pg no. 1-250.

2. Geary RC (1927) Some properties of correlation and regression in a limited universe, Metron 7: 83-119.

3. Eden T, Yates F (1993) On the validity of Fisher's z test when applied to an actual sample of non-normal data. J Agric Sci 23: 6-17.

4. Pitman EJG (1937) Significance tests which may be applied to samples from any populations Suppl J Royal Stat Soc 4: 119-130.

5. Pitman EJG (1937) Significance tests which may be applied to samples from any populations: II. The correlation coefficient test. Suppl J Royal Stat Soc 4: 225-232.

6. Pitman EJG (1938) Significance tests which may be applied to samples from any populations: III. The analysis of variance test, Biometrika 29: 322-335.

7. Johnston JE, Berry KJ, Mielke PW (2007) Permutation tests: Precision in estimating probability values. Percept Mot Skills 105 (3): 915-920.

8. KJ Berry, JE Johnston, Mielke PW (2014) A Chronicle of Permutation Statistical Methods, Springer-Verlag, Cham, USA.

9. Hilbert M (2012) How much information is there in the "information society"? Significance 9(4): 8-12.

10. Fisher RA (1935) The logic of inductive inference. J Royal Stat Soc 98(1): 39-82.

11. Lange J (1931) Crime as Destiny: A Study of Criminal Twins, Allen & Unwin, London.

12. Quetelet LAJ (1849) Letters Addressed to H.R.H. the Grand Duke of Saxe Coberg and Gotha, on the Theory of Probabilities as Applied to the Moral and Political Sciences, Charles & Edwin Layton, London, 1849.

13. Scheffe H (1959) The Analysis of Variance, Wiley, New York, Pp. 477.

14. Kendall MG (1938) A new measure of rank correlation Biometrika 30(1-2): 81-93.

15. Kendall MG (1948) Rank Correlation Methods, Griffin, London.

16. Kendall MG, Babington S (1939) The problem of m rankings. Ann Math Stat 10(3): 275-287.

17. Pearson K (1900) On the criterion that a given system of deviations from the probable in the case of acorrelated system of variables is such that it can be reasonably supposed to have arisen from random sampling, Philos Mag 50: 157-175.

18. Wilks SS (1935) The likelihood test of independence in contingency tables. Ann Math Stat 6(4): 190-196.

19. Wilks SS (1938) The large-sample distribution of the likelihood ratio for testing composite hypotheses. Ann Math Stat 9(1): 60-62.

20. Freeman MF, Tukey JW (1950) Transformations related to the angular and the square root, Ann Math Stat 21(4): 607-611.

21. Cressie N, Read TRC (1984) Multinomial goodness-of-fit tests. J Royal Stat Soc B 46:440-464.

22. Hardy GH, Ramanujan S(1918) Asymptotic formulae in combinatory analysis, Proc Lond Math Soc 17 (1): 75-115.

23. Mielke PW, Berry KJ (1993) Exact goodness-of-fit probability tests for analyzing categorical data. Educ Psychol Meas 53: 707-710.

24. Berry KJ, Johnston JE, Mielke PW (2004) Exact goodness-of-fit tests for unordered equiprobable Categories. Percept Mot Skills 98(3): 909-918.

25. Berry KJ, Mielke PW, Mielke HW (2002) The Fisher-Pitman permutation test: An attractive alternative to the F test. Psych Rep 90(2): 495-502.

26. Euler L (1988) Introduction to Analysis of the Infinite, Springer-Verlag, New York.

Your next submission with Juniper Publishers
will reach you the below assets

- Quality Editorial service
- Swift Peer Review
- Reprints availability
- E-prints Service
- Manuscript Podcast for convenient understanding
- Global attainment for your research
- Manuscript accessibility in different formats
  (Pdf, E-pub, Full Text, Audio)
- Unceasing customer service

Track the below URL for one-step submission
https://juniperpublishers.com/online-submission.php