Chapter 3
Permutation Statistical Methods

Abstract This chapter presents two models of statistical inference: the conventional Neyman–Pearson population model that is taught in every introductory course and the Fisher–Pitman permutation model with which the reader is assumed to unfamiliar. The Fisher–Pitman model consists of three different permutation methods: exact permutation methods, Monte Carlo permutation methods, and moment-approximation permutation methods. The three methods are described and illustrated with example analyses.

This chapter presents two competing models of statistical inference: the population (normal) model and the permutation model. The Neyman–Pearson population model is the standard model taught in all introductory classes and is familiar to most readers.[1] The Neyman–Pearson population model was specifically designed to make inferences about population parameters, provide approximate probability values, and is characterized by the assumptions of random sampling, a normally-distributed population, and homogeneity of variance when appropriate. The Fisher–Pitman permutation model of statistical inference is less well known and includes three different permutation methodologies, each of which is described and illustrated in this chapter: exact permutation methods, Monte Carlo permutation methods, and moment-approximation permutation methods.[2] In contrast to conventional statistical tests based on the Neyman–Pearson population model, tests based on the Fisher–Pitman permutation model are distribution-free, entirely data-dependent, appropriate for nonrandom samples, provide exact probability values, and are ideal for small sets of data.

[1]The Neyman–Pearson population model of statistical inference is named for Jerzy Neyman (1894–1981) and Egon Pearson (1895–1980).

[2]The Fisher–Pitman permutation model of statistical inference is named for R.A. Fisher (1890–1962) and E.J.G. Pitman (1897–1993).

K. J. Berry et al., *A Primer of Permutation Statistical Methods*,
https://doi.org/10.1007/978-3-030-20933-9_3

On the other hand, permutation tests can be computationally intensive, often requiring many millions of calculations. Five computational efficiencies for permutation statistical tests are described in this chapter. First, the development of high-speed computing has made permutation methods feasible. Second, the examination of all combinations of the observed data instead of all permutations of the data greatly reduces the amount of calculation required. Third, the use of mathematical recursion simplifies calculations of both test statistics and probability values. Fourth, calculation of only the variable portion of the selected test statistic minimizes the calculations required. Fifth, holding one array of the observed data constant reduces the number of arrangements required for exact permutation analyses.

As documented in Chap. 2, the permutation model of statistical inference had its beginnings in the 1920s and 1930s with the works of Fisher [12], Geary [14], Eden and Yates [9], Hotelling and Pabst [18], and Pitman [36–38]. Constrained by the difficulty of computing tens of thousands of statistical values on tens of thousands of arrangements of the observed data, permutation methods languished for many years until the advent of high-speed computing. Presently, statistical methods under the Fisher–Pitman permutation model is a rapidly developing field of statistical methodology and finds increasing utility in a large number of academic fields and disciplines.

3.1 The Neyman–Pearson Population Model

In contemporary research two competing models of statistical inference coexist: the population model and the permutation model.[3] The population model of statistical inference, formally proposed by Jerzy Neyman and Egon Pearson in a seminal two-part article on statistical inference published in 1928, is the model taught almost exclusively in introductory courses, although in most textbooks the presentation of the population model espoused by Neyman and Pearson is often conflated with an approach espoused by Fisher [19].

The Neyman–Pearson population model of statistical inference assumes random sampling with replacement from one or more specified populations [34, 35]. Under the Neyman–Pearson population model the level of statistical significance that results from applying a statistical test to the results of an experiment or survey corresponds to the frequency with which the null hypothesis would be rejected in repeated random samplings from the same specified population(s). Because repeated sampling of the specified population(s) is usually prohibitive, it is assumed that an approximating theoretical distribution such as a z, t, F, or χ^2 distribution conforms

[3]There are, of course, other models of statistical inference. A third model, the Bayesian inference model, is also very popular, especially in the decision-making sciences.

to the discrete sampling distribution of the test statistics generated under repeated random sampling.

Under the Neyman–Pearson population model two hypotheses concerning a population parameter or parameters are advanced: the null hypothesis symbolized by H_0 and a mutually-exclusive, exhaustive alternative hypothesis symbolized by H_1.[4] The probability of rejecting a true H_0 is determined by the researcher and specified as type I or α error, a region of rejection in the tail or tails of the theoretical distribution is delimited corresponding to α; for example, $\alpha = 0.05$ or $\alpha = 0.01$, and H_0 is rejected if the observed test statistic value falls into the region(s) of rejection with probability of type I error equal to or less than α.

Technically, under the Neyman–Pearson population model of statistical inference the null hypothesis is rejected if the computed test statistic value falls into the region of rejection defined by α. For example, if $\alpha = 0.05$ with a two-tail test and the critical values defining the region of rejection are ± 1.96, then a test statistic value more extreme than ± 1.96 in either direction implies rejection of the null hypothesis with a probability of type I error usually expressed as $p < 0.05$. In this research monograph asymptotic probability values under the Neyman–Pearson population model are given to four decimal places for comparison with exact probability values under the Fisher–Pitman permutation model of statistical inference.

3.2 The Fisher–Pitman Permutation Model

While the Neyman–Pearson population model of statistical inference is familiar to most researchers, the Fisher–Pitman permutation model of inference may be less familiar. Permutation statistical methods were introduced by R.A. Fisher in 1925 [12], further developed by Geary in 1927 [14], Eden and Yates in 1933 [9], Hotelling and Pabst in 1936 [18], and made explicit by Pitman in 1937 and 1938 [36–38]. For the interested reader, a number of excellent presentations of the two models are available. See especially, discussions by Curran-Everett [8], Feinstein [11], Hubbard [19], Kempthorne [23], Kennedy [24], Lachin [25], Ludbrook [26, 27], and May and Hunter [30].

For a permutation statistical test in its most basic form, a test statistic is computed on the observed data—often the same test statistic as in the Neyman–Pearson population model. The observations are then permuted over all possible arrangements of the observed data and the specified statistic is computed for each possible, equally-likely arrangement of the observed data. The proportion of arrangements in the reference set of all possible arrangements possessing test statistic values that are equal to or more extreme than the observed test statistic value constitutes the probability of the observed test statistic value.

[4] Some introductory textbooks denote the alternative hypothesis by H_A.

Figure 3.1 presents a flowchart detailing the calculation of an exact permutation probability value under the Fisher–Pitman model. The first step is to initialize two counters; in this case, Counter A and Counter B. Counter A provides a count of all test statistic values that are equal to or greater than the observed test statistic value. Counter B provides a count of all possible arrangements of the observed data. Second, the desired test statistic is calculated on the observed set of data. Third, a new arrangement of the observed data is generated, while preserving the sample size(s) and Counter B is increased by 1. Fourth, the desired test statistic is calculated on the new arrangement of the observed data and compared with the original test statistic value calculated on the observed set of data. If the value of the new test statistic is equal to or greater than the value of the observed test statistic, Counter A is increased by 1. If not, a check is made to see if this arrangement is the last in the reference set of all possible arrangements. If it is, then Counter A divided by Counter B yields the exact probability value; that is, the proportion of all possible test statistic values that are equal to or greater than the observed test statistic value. Otherwise, a new arrangement of the observed data is generated and the process is repeated.

Statistical tests and measures based on the Fisher–Pitman permutation model possess several advantages over statistical tests and measures based on the Neyman–Pearson population model. First, tests based on the permutation model are much less complex than tests based on the population model. Therefore, the results are much easier to communicate to unsophisticated or statistically naïve audiences. Second, permutation tests provide exact probability values based on the discrete permutation distribution of equally-likely test statistic values. Tests based on the Neyman–Pearson population model only provide vague results such as $P < 0.05$.[5] Third, permutation tests are entirely data-dependent in that all the information required for analysis is contained within the observed data—also called "the data at hand method" [16]. There is no reliance on factors external to the observed data, such as population parameters, assumptions about theoretical approximating distributions, and alternative hypotheses. Fourth, permutation tests are appropriate for nonrandom samples, such as are common in many fields of research. Fifth, permutation tests are distribution-free in that they do not depend on the assumptions associated with conventional tests under the population model, such as normality and homogeneity of variance. Sixth, permutation tests are ideal for small data sets, where conventional tests often are problematic when attempting to fit a continuous theoretical distribution to only a few discrete values.

Because permutation statistical methods are inherently computationally-intensive, it took the development of high-speed computing for permutation methods to achieve their potential. Today, a small laptop computer outperforms even the largest mainframe computers of previous decades. Three types of permutation tests are common in the literature: exact, Monte Carlo, and moment-approximation permutation tests.

[5] In this book, an upper-case letter P indicates a cumulative probability value and a lower-case letter p indicates a point probability value.

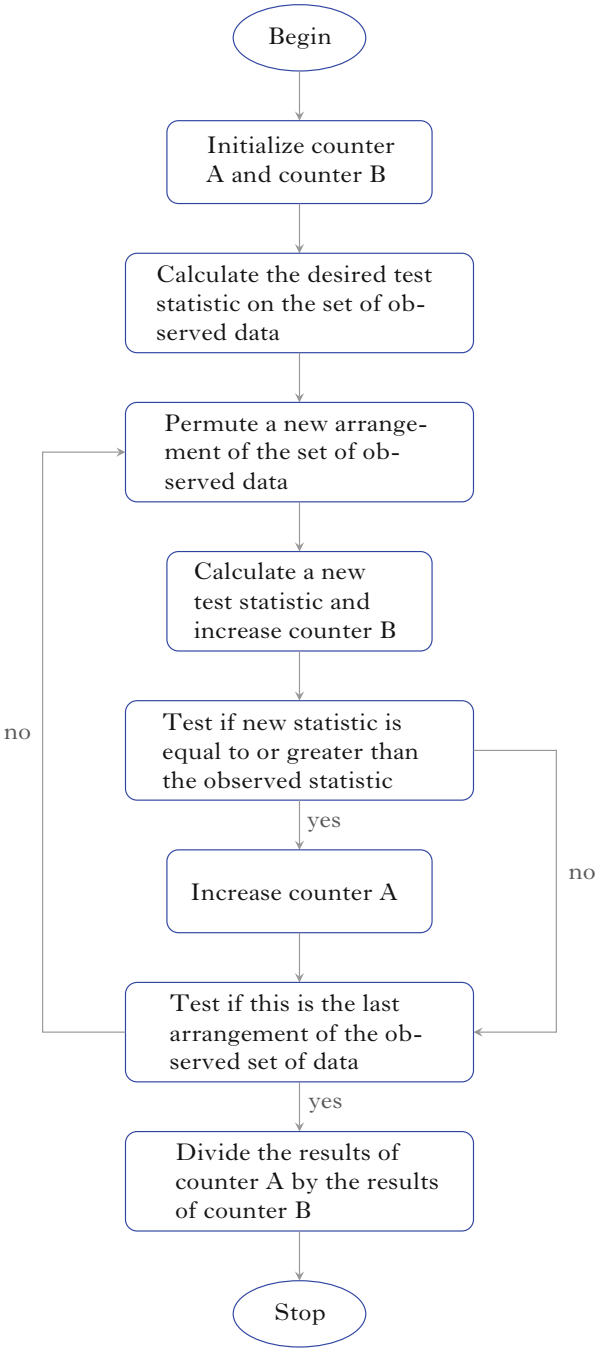

Fig. 3.1 Flowchart for the calculation of an exact permutation probability value

Table 3.1
Cross-classification of variables A and B

Variable A	Variable B		Total
	b	$\bar{b}$	
a	9	9	18
$\bar{a}$	0	12	12
Total	9	21	30

3.2.1 Exact Permutation Tests

The first step in an exact permutation test is to calculate a test statistic value for the observed data. Second, a reference set of all possible, equally-likely arrangements of the observed data is systematically generated. Third, the desired test statistic is calculated for each arrangement in the reference set. Fourth, the probability of obtaining the observed value of the test statistic, or one more extreme, is the proportion of the test statistics in the reference set with values that are equal to or more extreme than the value of the observed test statistic.

To be perfectly clear, in practice a different order is followed. First, a test statistic value for the observed data is calculated. Second, the first of a reference set of all possible, equally-likely arrangements of the observed data is generated. Third, a test statistic value for the new arrangement of the observed data is calculated and compared with the original test statistic value. Fourth, if the new value is equal to or exceeds the original test statistic value, a counter is increased by one. The process is repeated until all possible arrangements of the observed data have been generated and evaluated. Finally, the probability of obtaining the observed value of the test statistic, or one more extreme, is the proportion of the test statistics in the reference set with values that are equal to or more extreme than the value of the observed test statistic. In this manner it is not necessary to store the reference set of all possible arrangements of the observed data, which is often quite large.

An Exact Permutation Example

To illustrate an exact permutation test, consider the small set of data given in Table 3.1. Fisher's exact probability test is the iconic permutation test.[6] Fisher's exact test calculates the hypergeometric point probability value for the reference set of all possible arrangements of cell frequencies, given the observed marginal frequency totals. The two-tail probability value of the observed arrangement of cell frequencies is the sum of the observed probability value and all probability values that are equal to or less than the observed probability value. Because Fisher's exact test simply yields a probability value, there is no test statistic defined in the

[6]Fisher's exact test was independently developed by R.A. Fisher, Joseph Irwin, and Frank Yates in the early 1930s [13, 21, 40].

Table 3.2 Conventional notation for a 2×2 contingency table

Variable A	Variable B		Total
	b	$\bar{b}$	
a	n_{11}	n_{12}	$n_{1.}$
$\bar{a}$	n_{21}	n_{22}	$n_{2.}$
Total	$n_{.1}$	$n_{.2}$	N

usual sense. Thus the first step is to determine the reference set of all possible arrangements of the four cell frequencies, given the observed marginal frequency totals. For a 2×2 contingency table, it is relatively easy to determine the total number of possible tables in the reference set.

Consider the 2×2 contingency table in Table 3.2. Denote by a dot (·) 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, and $n_{.j}$ denotes the marginal frequency total of the jth column, $j = 1, \ldots, c$, summed over all rows. Thus $n_{1.}$ and $n_{2.}$ denote the marginal frequency totals for rows 1 and 2, $n_{.1}$ and $n_{.2}$ denote the marginal frequency totals for columns 1 and 2, n_{ij} denotes the cell frequencies for $i, j = 1, 2$, and $N = n_{11} + n_{12} + n_{21} + n_{22}$. The total number of possible values for any cell frequency, say n_{11}, is given by

$$M = \min(n_{1.}, n_{.1}) - \max(0, n_{11} - n_{22}) + 1 \, .$$

Thus, for the frequency data given in Table 3.1 there are

$$M = \min(18, 9) - \max(0, 8 - 11) + 1 = 9 - 0 + 1 = 10$$

possible arrangements of cell frequencies in the reference set, given the observed row and column marginal frequency distributions, {18, 12} and {9, 21}, respectively.

The reference set of the $M = 10$ arrangements of cell frequencies and the associated hypergeometric point probability values are listed in Table 3.3. For any 2×2 contingency table, such as depicted in Table 3.2, the hypergeometric point probability of any specified cell, say cell (1,1), is given by

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

For the frequency data given in Table 3.1, the two-tail probability value is the sum of the probability value of the observed contingency table and all probability values that are equal to or less than the probability value of the observed table. Thus Table 10 in Table 3.3 (the observed table) has a hypergeometric point probability value of $p_{10} = 0.3398 \times 10^{-2}$ and only Tables 3.1 and 3.2 possess point probability values that are less than $p = 0.3398 \times 10^{-2}$; that is, $p_1 = 0.1538 \times 10^{-4}$ and

Table 3.3 Listing of the $M = 10$ possible 2×2 contingency tables in the reference set from the frequency data given in Table 3.1 with associated exact hypergeometric point probability values

Table 1		Probability	Table 2		Probability
0	18	0.1538×10^{-4}	1	17	0.6228×10^{-3}
9	3		8	4	
Table 3		Probability	Table 4		Probability
2	16	0.8470×10^{-2}	3	15	0.5270×10^{-1}
7	5		6	6	
Table 5		Probability	Table 6		Probability
4	14	0.1694×10^{-1}	5	13	0.2964
5	7		7	8	
Table 7		Probability	Table 8		Probability
6	12	0.2855	7	11	0.1468
3	9		5	10	
Table 9		Probability	Table 10		Probability
8	10	0.3670×10^{-1}	9	9	0.3398×10^{-2}
1	11		0	12	

Table 3.4 Listing of the 3×5 cell frequencies with rows (R_1, R_2, R_3) and columns $(C_1, C_2, C_3, C_4, C_5)$ for an exact probability example

	B_1	B_2	B_3	B_4	B_5	Total
A_1	4	7	2	9	0	22
A_2	1	5	2	7	6	21
A_3	4	5	10	18	0	37
Total	9	17	14	34	6	80

$p_2 = 0.6228\times10^{-3}$, respectively. The cumulative probability value of the three tables is

$$\begin{aligned} P &= p\{9|18, 9, 30\} + p\{0|18, 9, 30\} + p\{1|18, 9, 30\} \\ &= \frac{18!\,12!\,9!\,21!}{30!\,9!\,9!\,0!\,12!} + \frac{18!\,12!\,9!\,21!}{30!\,0!\,18!\,9!\,3!} + \frac{18!\,12!\,9!\,21!}{30!\,1!\,17!\,8!\,4!} \\ &= 0.3398\times10^{-2} + 0.1538\times10^{-4} + 0.6228\times10^{-3} \\ &= 0.4036\times10^{-2}\,. \end{aligned}$$

A Second Exact Permutation Test Example

For a second example of an exact permutation analysis, consider the 3×5 contingency table with $N = 80$ cell frequencies given in Table 3.4. Pearson's chi-squared test statistic for an $r\times c$ contingency table is taught in every introductory course and

is given by

$$\chi^2 = N\left(\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{n_{ij}^2}{n_{i.}n_{.j}} - 1\right),$$

where $n_{i.}$ denotes a row marginal frequency total for $i = 1, \ldots, r$, $n_{.j}$ denotes a column marginal frequency total for $j = 1, \ldots, c$, n_{ij} denotes an observed cell frequency for $i = 1, \ldots, r$ and $j = 1, \ldots, c$, and N is the total of the cell frequencies; in this case, $N = 80$. For the frequency data given in Table 3.4 with row marginal frequency totals $\{22, 21, 37\}$ and column marginal frequency totals $\{9, 17, 14, 34, 6\}$, the observed value of Pearson's chi-squared test statistic is

$$\begin{aligned}\chi^2 &= N\left(\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{n_{ij}^2}{n_{i.}n_{.j}} - 1\right)\\ &= 80\left(\frac{4^2}{(22)(9)} + \frac{7^2}{(22)(17)} + \cdots + \frac{0^2}{(37)(6)} - 1\right) = 25.1809\,.\end{aligned}$$

The exact probability value of $\chi^2 = 25.1809$ under the Fisher–Pitman permutation model is the sum of the hypergeometric point probability values associated with the chi-squared values calculated on the reference set of all M possible arrangements of the cell frequencies, given the observed marginal frequency totals. For the frequency data given in Table 3.4, there are $M = 21{,}671{,}722$ possible, equally-likely arrangements of the cell frequencies given the observed marginal frequency totals, of which 16,498,422 chi-squared values are equal to or greater than the observed chi-squared value of $\chi^2 = 25.1809$, yielding an exact hypergeometric probability value of $P = 0.1009 \times 10^{-2}$.

For comparison, the chi-squared test statistic is asymptotically distributed as Pearson's χ^2 with $(r-1)(c-1)$ degrees of freedom under the Neyman–Pearson null hypothesis. With $(r-1)(c-1) = (3-1)(5-1) = 8$ degrees of freedom, the asymptotic probability value of $\chi^2 = 25.1809$ is $P = 0.1449 \times 10^{-2}$.

Comparison with Fisher's Exact Probability Test

Although Fisher's exact probability test is typically limited to 2×2 contingency tables, it is possible to compute Fisher's exact test on larger tables, such as the 3×5 contingency table given in Table 3.4 [32]. It is important to note that Fisher's exact probability test and an exact chi-squared test of independence are constructed quite differently, although both tests will occasionally yield identical probability values.

Fisher's exact test generates a reference set of all M possible arrangements of cell frequencies given the observed marginal frequency totals, computes the hypergeometric point probability value for each arrangement of the observede data, and sums

the probability values that are equal to or less than the probability value obtained from the observed arrangement of cell frequencies. On the other hand, an exact chi-squared test generates a reference set of all M possible arrangements of cell frequencies given th observed marginal frequency totals, calculates the chi-squared value for each arrangement of cell frequencies, computes the hypergeometric point probability value for each arrangement, and sums the probability values associated with those chi-squared values that are equal to or greater than the chi-squared value obtained from the observed arrangement of cell frequencies.

For the frequency data given in Table 3.4, the point probability value for the observed arrangement of cell frequencies is $p = 0.5164 \times 10^{-8}$. There are $M = 21{,}671{,}722$ possible, equally-likely arrangements of the cell frequencies in Table 3.4, of which 18,683,509 hypergeometric point probability values are equal to or greater than $p = 0.5164 \times 10^{-8}$, yielding an exact probability value of $P = 0.5174 \times 10^{-2}$.

3.2.2 Monte Carlo Permutation Tests

As sample sizes increase, the size of the reference set of all possible arrangements of the observed data can become quite large and exact permutation methods are quickly rendered impractical. For example, permuting two samples of sizes $n_1 = n_2 = 35$ generates

$$M = \frac{(n_1 + n_2)!}{n_1!\, n_2!} = \frac{(35 + 35)!}{35!\, 35!} = 112{,}186{,}277{,}816{,}662{,}845{,}432$$

equally-likely arrangements of the observed data; or in words, 112 billion billion different arrangements of the observed data—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 substituted, providing approximate, but highly accurate, probability values. Monte Carlo permutation methods generate and examine a random subset of all possible, equally-likely arrangements of the observed data. 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 probability values that are equal to or more extreme than the probability value of the observed test statistic. With a sufficient number of randomly-selected samples, a probability value can be computed to any reasonable accuracy. Provided the probability value is not too small, the current recommended practice is to use $L = 1{,}000{,}000$ randomly-selected arrangements of the observed data to ensure a probability value with three decimal places of accuracy. To ensure four decimal places of accuracy, the number of randomly-selected arrangements must be increased by two magnitudes of order; that is, $L = 100{,}000{,}000$ [22].

A Monte Carlo Permutation Example

Consider once again the frequency data given in Table 3.4 on p. 64 with $N = 80$ observations. In many cases the exact analysis of $M = 21{,}671{,}722$ arrangements of cell frequencies would be considered impractical. In such cases a random sample of cell arrangements can yield an approximate probability value with considerable accuracy. Based on $L = 1{,}000{,}000$ randomly-selected cell arrangements given the observed marginal frequency totals, the Monte Carlo probability value of $\chi^2 = 25.1809$ is $P = 0.1055 \times 10^{-2}$, which compares favorably with the exact probability value of $P = 0.1009 \times 10^{-2}$.

3.2.3 Moment-Approximation Permutation Tests

Monte Carlo permutation methods can be inefficient when desired probability values are very small; for example, probability values on the order of 10^{-6}, as the Monte Carlo permutation method requires a large number of randomly-selected test statistics to approximate such a small probability value. Prior to the development of high-speed computing that made exact and Monte Carlo permutation methods possible, researchers relied on moment-approximation procedures to provide approximate probability values. The moment-approximation of a test statistic requires calculation of the exact moments of the test statistic, assuming equally-likely arrangements of the observed data. The exact moments are then used to fit a specified distribution that approximates the underlying discrete permutation distribution and provide an approximate, but often highly accurate, probability value.

For many years the beta distribution was used for the approximating distribution. Presently, the approximating distribution of choice is the Pearson type III probability distribution, which depends on the exact mean, variance, and skewness of the test statistic under consideration, say δ, given by

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

$$\sigma_\delta^2 = \frac{1}{M} \sum_{i=1}^{M} \left(\delta_i - \mu_\delta\right)^2 \ ,$$

and

$$\gamma_\delta = \frac{1}{\sigma_\delta^3} \left[\frac{1}{M} \sum_{i=1}^{M} \left(\delta_i - \mu_\delta\right)^3 \right] \ ,$$

respectively, where M denotes the total number of possible, equally-likely arrangements of the observed data. The standardized statistic given by

$$T = \frac{\delta_o - \mu_\delta}{\sigma_\delta}$$

follows the Pearson type III distribution, where δ_o denotes the observed value of test statistic δ. It should be noted that while the moments are exact, the resultant Pearson type III probability value is always approximate.

A Moment-Approximation Permutation Example

For the frequency data given in Table 3.4 on p. 64, the observed value of the permutation test statistic is $\delta_o = 24.8661$, the expected value of test statistic δ is $\mu_\delta = 8.00$, the variance of test statistic δ is $\sigma_\delta^2 = 14.5148$, the standardized test statistic is

$$T = \frac{\delta_o - \mu_\delta}{\sigma_\delta} = \frac{24.8661 - 8.00}{\sqrt{14.5148}} = +4.4270\,,$$

and the moment-approximation probability value based on the Pearson type III probability distribution is $P = 0.9763{\times}10^{-3}$.

A Comparison of the Three Approaches

The three approaches to determining permutation probability values (exact, Monte Carlo, and moment-approximation) often yield similar probability values. The difference between the moment-approximation probability value ($P = 0.9763{\times}10^{-3}$) and the exact probability value based on all $M = 21{,}671{,}722$ arrangements of the observed data in Table 3.4 ($P = 0.1009{\times}10^{-4}$) is only

$$\Delta_P = 0.9763{\times}10^{-3} - 0.1009{\times}10^{-4} = 0.9662{\times}10^{-3}\,,$$

the difference between the moment-approximation probability value ($P = 0.9763{\times}10^{-3}$) and the Monte Carlo probability value based on a sample of $L = 1{,}000{,}000$ random arrangements of the observed data in Table 3.4 ($P = 0.1055{\times}10^{-2}$) is only

$$\Delta_P = 0.1055{\times}10^{-2} - 0.9763{\times}10^{-3} = 0.7870{\times}10^{-4}\,,$$

and the difference between the Monte Carlo probability value based on a sample of $L = 1{,}000{,}000$ random arrangements of the observed data in Table 3.4 ($P = 0.1055{\times}10^{-2}$) and the exact probability value based on all $M = 21{,}671{,}722$ arrangements of the observed data in Table 3.4 ($P = 0.1009{\times}10^{-2}$) is only

$$\Delta_P = 0.1055{\times}10^{-2} - 0.1009{\times}10^{-2} = 0.4600{\times}10^{-4}\,.$$

3.3 Permutation and Parametric Statistical Tests

Permutation statistical tests, based on the Fisher–Pitman permutation model, differ from traditional parametric tests, based on the Neyman–Pearson population model, in several ways. First, permutation tests are entirely data-dependent in that all the information required for analysis is contained within the observed data set [4, 33]. Second, permutation tests are appropriate for nonrandom samples, such as are common in many fields of research [38]. Third, permutation tests are distribution-free in that they do not depend on the assumptions associated with traditional parametric tests, such as normality and homogeneity of variance [5]. Fourth, permutation tests provide exact probability values based on the discrete permutation distribution of equally-likely test statistic values, rather than approximate probability values based on a theoretical approximating distribution, such as a z, χ^2, t, or F distribution [11]. Fifth, permutation tests are ideal for small data sets, whereas distribution functions often provide poor fits to the underlying discrete sampling distribution. Of these five differences, the requirements of random sampling and normality greatly limit the application of statistical tests and measures based on the Neyman–Pearson population model.

3.3.1 The Assumption of Random Sampling

It is important to note that the mathematical theorems that justify most statistical procedures under the Neyman–Pearson population model of statistical inference apply only to random samples drawn with replacement from a completely-specified sampling frame. However, if the sample is not a random sample from a well-defined population, then the validity of the hypothesis test is questionable [38]. There are, admittedly, some applications in statistical analysis in which random sampling from a specified population is neither attempted nor considered important. The fact that medical researchers seldom use random samples often comes as a surprise to investigators who work in other domains [11].

Research psychologists have been especially concerned with problems of random sampling. Writing in *Psychological Bulletin* in 1966, psychologist Eugene Edgington stated his position unequivocally: "statistical inferences cannot be made concerning populations that have not been randomly sampled" [10, p. 485]. Writing in *Canadian Psychology* in 1993, psychologists Michael Hunter and Richard May noted that random sampling is of particular relevance to psychologists, "who rarely use random sampling or any other sort of probability sampling" [20, p. 385]. In 1988 psychologist William Hays wrote:

> The point is that *some* probability structure must be known or assumed to underlie the occurrence of samples if statistical inference is to proceed. This point is belabored only because it is so often overlooked, and statistical inferences are so often made with only the most casual attention to the process by which the sample was generated. The assumption

> of some probability structure underlying the sampling is a little "price tag" attached to a statistical inference. It is a sad fact that if one knows nothing about the probability of occurrence for particular samples of units for observation, very little of the machinery we are describing here applies. This is why our assumption of random sampling is not to be taken lightly. . . . Unless this assumption is at least reasonable, the probability results of inferential methods mean very little, and these methods might as well be omitted [17, p. 212].[7]

In summary, conventional sampling distributions require random sampling whereas permutation distributions do not [20, p. 387].

3.3.2 The Assumption of Normality

The assumption of normality is so basic to classical statistics that it deserves special attention. Two points should be emphasized. First, permutation tests make no distributional assumptions and, therefore, do not depend on the assumption of normality. Second, the assumption of normality by conventional tests is always unrealistic and never justified in practice [5, 29].

In 1927 R.C. Geary famously proclaimed: "Normality is a myth; there never has, and never will be, a normal distribution" [15, p. 241] and in 1938 Joseph Berkson wrote: "we may assume that it is practically certain that any series of real observations does not actually follow a normal curve *with absolute exactitude* in all respects" [2, p. 526] (see footnote 7). Robert Matthews once described the normal distribution as "beautiful, beguiling and thoroughly dangerous" [29, p. 193] and in 1954 I.D.J. Bross pointed out that statistical methods "are based on certain assumptions—assumptions which not only can be wrong, but in many situations *are* wrong" [6, p. 815] (see footnote 7). Others have empirically demonstrated the prevalence of highly-skewed and heavy-tailed distributions in a variety of academic disciplines, the best-known of which is Theodore Micceri's widely quoted 1989 article on "The unicorn, the normal curve, and other improbable creatures" [31].

3.4 Advantages of Permutation Methods

Alvan Feinstein was a strong advocate for permutation methods. Trained as both a mathematician and a medical doctor, Feinstein is widely regarded as the founder of clinical epidemiology and patient-oriented medicine and the originator of clinimetrics: the application of mathematics to the field of medicine [3, p. 246]. In 1973 Feinstein published a formative article titled "The role of randomization in sampling, testing, allocation, and credulous idolatry" [11]. As Feinstein's focus

[7] Emphasis in the original.

was on medical investigations, he detailed the major violations of the assumptions underlying tests of two groups:

1. The groups studied in modern clinical or epidemiologic research are seldom selected as random samples.
2. For the many clinical and epidemiology research projects that are performed as surveys, the subjects are not randomly assigned.
3. The distribution of the target variable is usually unknown in the parent population.
4. It is usually known that the target variable does not have a Gaussian distribution, and often departs from it dramatically.
5. It is usually known that the variances of the two samples are not remotely similar.

Feinstein then put forth some advantages of tests under the Fisher–Pitman permutation model that were insightful for the time and foreshadowed later research:

1. The result of a permutation test is a direct, exact probability value for the random likelihood of the observed difference.
2. Permutation tests do not require any unwarranted inferential estimations of means, variances, pooled variances, or other parameters of an unobserved, hypothetical parent population. The tests are based solely on the evidence that was actually obtained.
3. The investigator is not forced into making any erroneous assumptions either that the contrasted groups were chosen as random samples from a parent population or that treatments under study were randomly allocated to the two groups.
4. The investigator is not forced into making any erroneous or unconfirmable assumptions about a Gaussian (or any other) distribution for the parent population, or about equal variances in the contrasted groups.
5. A permutation test can be applied to groups of any size, no matter how large or small. There are no degrees of freedom to be considered. In the case of a contingency table, there is no need to worry about the magnitude of the expected value, no need to calculate expectations based on fractions of people, and no need to worry about applying, or not applying, Yates' correction for continuity.

To summarize, permutation statistical methods yield exact probability values, are completely data-dependent, do not require random sampling, make no assumptions about distributions, and can be applied to very small samples. The one drawback to permutation tests, as noted by Feinstein in 1973, is that permutation tests are notoriously difficult to calculate. While this statement was certainly true in 1973, in the age of high-speed computing the statement is most certainly no longer accurate.

3.5 Calculation Efficiency

Although permutation statistical methods do not require random sampling, normality, homogeneity, or large sample sizes, a potential drawback is the sheer amount of computation required, with exact permutation tests being unrealistic for many statistical analyses. Even Monte Carlo permutation methods often require the enumeration of millions of random arrangements of the observed data in order to provide a desired accuracy.

Five innovations mitigate the computation problem. First, high-speed computing makes possible exact permutation statistical methods in which all possible arrangements of the observed data are generated and examined. Second, the examination of all combinations of the observed data instead of all permutations of the observed data provides the same probability value with considerable savings in computing time. Third, mathematical recursion greatly simplifies difficult calculations. Fourth, calculation of only the variable components of the selected test statistic reduces the amount of calculation required for each of the enumerated arrangements. Fifth, holding one array of the observed data constant in any type of block design can substantially lessen the number of arrangements required for an exact permutation analysis.

3.5.1 High-Speed Computing

One has only to observe the hordes of the digitally distracted trying to navigate a crowded sidewalk with their various smart-phones, pads, pods, ear-buds, and tablets to realize that computing power, speed, and accessibility have finally arrived. Permutation methods are, by their very nature, computationally intensive and required the development of high-speed computing to achieve their potential. Prior to 1960, computers were large, slow, and expensive. In large part their use was restricted to military and industrial applications. In the 1960s, mainframe computers became widely available to academicians at major research universities. By 1980 desktop computers and workstations, although not common, were available to many researchers. In addition, the speed of computing increased greatly between 1960 and 1980. All this paved the way for the rapid development of permutation statistical 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 1905 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[8]; 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% with the introduction of China's Tianhe-1A supercomputer performing at 2.67 petaflops [28].

In October of 2011, China broke the petaflops barrier again with the introduction of the Sunway Bluelight MPP [1]. 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. In June of 2018 IBM unveiled the Summit supercomputer at Oak Ridge National Laboratory in Tennessee that achieved sustained computing speeds of 200 petaflops.

On the near horizon are so-called quantum computers. The basic element of a quantum computer is the qubit. Unlike a standard bit (binary digit), which can take on a value of either 0 or 1, a qubit (quantum bit) can be either 0, 1, or a combination of the two. Because qubits can represent 0 and 1 simultaneously, they can encode a wealth of information. As Thomas Siegfried explained it, five bits represent *one* out of $2^5 = 32$ possible permutations, but five qubits represent *all* of $2^5 = 32$ possible permutations [39]. Teams from academia and industry are working on versions of quantum computers with 50–100 qubits, enough to perform calculations that the most powerful supercomputers of today cannot accomplish in a reasonable time [7]. Google, which has already developed a nine qubit computer, has aggressive plans to scale up to 49 qubits, and IBM, which has developed a 16 qubit prototype, announced in early 2017 that it would build a 50 qubit quantum computer in the next few years [7].

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 permutation statistical methods. Combined with other efficiencies, it can safely be said that permutation methods have the potential to provide exact or Monte Carlo probability values in an efficient manner for a wide variety of statistical applications.

[8]One petaflops indicates a quadrillion operations per second, or a 1 with 15 zeroes following it.

3.5.2 Analysis with Combinations

Although permutation statistical methods are known by the attribution "permutation," they are generally not based on all possible permutations of the observed data. Instead, exact permutation methods are based on all possible *combinations* of arrangements of the observed data. Since, in general, there are fewer combinations than permutations, analysis of combinations of the observed data greatly reduces the amount of calculation required.

To illustrate the efficiency achieved by analyzing all combinations of the observed data instead of all permutations, consider $N = 10$ observations that are to be randomized into two groups, A and B, where $n_A = n_B = 5$ observations. Suppose that the purpose is to compare differences between the two groups, such as a mean difference. Let the $n_A = 5$ observations be designated as a through e and the $n_B = 5$ observations be designated as f through j. For Group A, the first observation can be chosen in 10 different ways, the second observation in nine ways, the third observation in eight ways, the fourth observation in seven ways, and the fifth observation in six ways. Once the five observations of Group A have been chosen, the remaining five observations are assigned to Group B.

Of the $10{\times}9{\times}8{\times}7{\times}6 = 30{,}240$ ways in which the five observations can be arranged for Group A, each individual quintet of observations will appear in a series of permutations. Thus, the quintet $\{a, b, c, d, e\}$ can be permuted as $\{a, b, c, e, d\}$, $\{a, b, d, e, c\}$, $\{a, b, d, c, e\}$, and so on. Each permutation of the five observations will yield the same mean value. The number of different permutations for a group of five observations is $5! = 120$. Thus, each distinctive quintet will appear in 120 ways among the 30,240 possible arrangements. Therefore, 30,240 divided by 120 yields 252 distinctive quintets of observations that can be formed by dividing $N = 10$ observations into two groups of five observations each. The number of quintets can conveniently be expressed as

$$\frac{(n_A + n_B)!}{n_A!\, n_B!} = \frac{(5+5)!}{5!\, 5!} = 252\ .$$

However, half of these arrangements are similar, but opposite. Thus, a quintet such as $\{a, b, c, d, e\}$ might be assigned to Group A and the quintet $\{f, g, h, i, j\}$ might be assigned to Group B, or vice versa, yielding the same absolute mean difference. Consequently, there are only $252/2 = 126$ distinctly different pairs of quintets to be considered. A substantial amount of calculation can be eliminated by considering all possible combinations of arrangements of the observed data in place of all possible permutations with no loss of accuracy. Even in this small example, a reduction from 30,240 equally-likely arrangements of the observed data to 126 arrangements constitutes a substantial increase in efficiency.

3.5.3 Mathematical Recursion

Mathematical recursion is a process by 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. 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 3.5. Denote by a dot (·) 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 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

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

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 the cell containing n_{11} observations is sufficient.

If

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

Table 3.5 Conventional notation for a 2×2 contingency table

	Category		
Category	1	2	Total
1	n_{11}	n_{12}	$n_{1.}$
2	n_{21}	n_{22}	$n_{2.}$
Total	$n_{.1}$	$n_{.2}$	N

then solving for $f(n_{11})$ produces

$$f(n_{11}) = \frac{p\{n_{11}+1|n_{1.}, n_{.1}, N\}}{p\{n_{11}|n_{1.}, n_{.1}, N\}}$$
$$= \frac{n_{11}!\, n_{12}!\, n_{21}!\, n_{22}!}{(n_{11}+1)!\,(n_{12}-1)!\,(n_{21}-1)!\,(n_{22}+1)!}$$

and, after cancelling, yields

$$f(n_{11}) = \frac{n_{12}\, n_{21}}{(n_{11}+1)(n_{22}+1)}\,. \tag{3.1}$$

To illustrate mathematical recursion with an arbitrary initial value, consider the 2×2 contingency table given in Table 3.6 with $N = 24$ observations. For the cell containing $n_{11} = 6$ observations there are

$$M = \min(n_{1.}, n_{.1}) - \max(0, n_{11} - n_{22}) + 1$$
$$= \min(10, 8) - \max(0, 6 - 12) + 1 = 8 - 0 + 1 = 9$$

possible arrangements of cell frequencies, given the observed marginal frequency totals. Table 3.7 lists the reference set of the $M = 9$ cell arrangements along with the associated hypergeometric point probability values to six decimal places.

To illustrate the use of an arbitrary origin in a recursion procedure, consider Table 3.1 in Table 3.7 and set relative probability value $H\{n_{11} = 0|10, 8, 24\}$ to a small arbitrarily-chosen positive value, say 1.00. Thus, $H\{n_{11} = 0|10, 8, 24\} = 1.00$. Then, following Eq. (3.1), a recursion procedure produces

$$H\{n_{11} = 1|10, 8, 24\} = 1.000000 \times \frac{(10)(8)}{(0+1)(6+1)} = 11.428571\,,$$
$$H\{n_{11} = 2|10, 8, 24\} = 11.428571 \times \frac{(9)(7)}{(1+1)(7+1)} = 45.000000\,,$$
$$H\{n_{11} = 3|10, 8, 24\} = 45.000000 \times \frac{(8)(6)}{(2+1)(8+1)} = 80.000000\,,$$
$$H\{n_{11} = 4|10, 8, 24\} = 80.000000 \times \frac{(7)(5)}{(3+1)(9+1)} = 70.000000\,,$$

Table 3.6 Example data for a recursion process with an arbitrary initial value

	Variable B		
Variable A	b	$\bar{b}$	Total
a	6	4	10
$\bar{a}$	2	12	14
Total	8	16	24

Table 3.7 Listing of the $M = 9$ possible 2×2 contingency tables from Table 3.6 in the reference set with associated exact hypergeometric probability values to six decimal places

Table 1		Probability	Table 2		Probability
0	10	0.004083	1	9	0.046664
8	6		7	7	
Table 3		Probability	Table 4		Probability
2	8	0.183739	3	7	0.326648
6	8		5	9	
Table 5		Probability	Table 6		Probability
4	6	0.285817	5	5	0.124720
4	10		3	11	
Table 7		Probability	Table 8		Probability
6	4	0.025983	7	3	0.002284
2	12		1	13	
Table 9		Probability			
8	2	0.000061			
0	14				

$$H\{n_{11} = 5|10, 8, 24\} = 70.000000 \times \frac{(6)(4)}{(4+1)(10+1)} = 30.545455\,,$$

$$H\{n_{11} = 6|10, 8, 24\} = 30.545455 \times \frac{(5)(3)}{(5+1)(11+1)} = \ 6.363636\,,$$

$$H\{n_{11} = 7|10, 8, 24\} = \ 6.363636 \times \frac{(4)(2)}{(6+1)(12+1)} = \ 0.559441\,,$$

and

$$H\{n_{11} = 8|10, 8, 24\} = \ 0.559441 \times \frac{(3)(1)}{(7+1)(13+1)} = \ 0.014985\,,$$

for a total of

$$T = \sum_{i=0}^{8} H\{n_{11} = i|10, 8, 24\}$$

$$= 1.000000 + 11.428571 + \cdots + 0.014985 = 244.912088\,.$$

The desired exact point probability values are then obtained by dividing each relative probability value, $H\{n_{11}|n_{1\cdot}, n_{\cdot1}, N\}$, by the recursively-obtained total, T.

For example,

$$p\{n_{11} = 0|10, 8, 24\} = \frac{H_1}{T} = \frac{1.000000}{244.912088} = 0.004083\,,$$

$$p\{n_{11} = 1|10, 8, 24\} = \frac{H_2}{T} = \frac{11.428571}{244.912088} = 0.046664\,,$$

$$p\{n_{11} = 2|10, 8, 24\} = \frac{H_3}{T} = \frac{45.000000}{244.912088} = 0.183739\,,$$

$$p\{n_{11} = 3|10, 8, 24\} = \frac{H_4}{T} = \frac{80.000000}{244.912088} = 0.326648\,,$$

$$p\{n_{11} = 4|10, 8, 24\} = \frac{H_5}{T} = \frac{70.000000}{244.912088} = 0.285817\,,$$

$$p\{n_{11} = 5|10, 8, 24\} = \frac{H_6}{T} = \frac{30.545455}{244.912088} = 0.124720\,,$$

$$p\{n_{11} = 6|10, 8, 24\} = \frac{H_7}{T} = \frac{6.363636}{244.912088} = 0.025983\,,$$

$$p\{n_{11} = 7|10, 8, 24\} = \frac{H_8}{T} = \frac{0.559441}{244.912088} = 0.002284\,,$$

and

$$p\{n_{11} = 8|10, 8, 24\} = \frac{H_9}{T} = \frac{0.014985}{244.912088} = 0.000061\,.$$

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 $\max(n_{11}) - \min(n_{11}) + 1$ is large, the computational savings can be substantial.

3.5.4 Variable Components of a Test Statistic

Under permutation, only the variable components of the specified test statistic need to be calculated for each arrangement of the observed data. As this component is often a very small piece of the desired test statistic, calculations can often be reduced

by several factors. To illustrate, consider the raw-score expression for a conventional Pearson product-moment correlation coefficient between variables x and y given by

$$r_{xy} = \frac{\sum_{i=1}^{N} x_i y_i - \left(\sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i\right) \Big/ N}{\sqrt{\left[\sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i\right)^2 \Big/ N\right]\left[\sum_{i=1}^{N} y_i^2 - \left(\sum_{i=1}^{N} y_i\right)^2 \Big/ N\right]}}, \tag{3.2}$$

where N is the number of bivariate measurements. For Pearson's correlation coefficient given in Eq. (3.2)

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

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

3.5.5 *Holding an Array Constant*

In the special case of block designs, such as matched-pairs and randomized-blocks analysis of variance, it is possible to reduce the number of arrangements to be examined by holding one of the arrays (treatment values) constant, while permuting the other arrays. For example, given $g = 3$ treatments and $b = 10$ subjects (blocks) in each treatment, there are

$$M = (g!)^b = (3!)^{10} = 60{,}466{,}176$$

arrangements of the observed data to be considered. Holding one of the b sets of blocks constant, relative to the other $b - 1$ sets of blocks, there are

$$M = (g!)^{b-1} = (3!)^{10-1} = 10{,}077{,}696$$

arrangements of the observed data to be considered, a reduction of 50,388,480 arrangements, or 83%.

These five features, high-speed computing, mathematical recursion with an arbitrary initial value, computation of only the variable components of the test statistic under permutation, holding an array of the observed data constant, and

utilizing combinations instead of permutations, produce a highly efficient permutation statistical approach that makes permutation statistical methods both feasible and practical for many research applications.

3.6 Summary

This chapter opened with a description of two models of statistical inference: the population model first put forward by Jerzy Neyman and Egon Pearson in 1928 and the permutation model developed by R.A. Fisher, R.C. Geary, T. Eden, F. Yates, H. Hotelling, M.R. Pabst, and E.J.G. Pitman in the 1920s and 1930s. Three types of permutation statistical methods were described and illustrated: exact, Monte Carlo, and moment-approximation permutation methods.

Permutation statistical methods were shown to differ from traditional parametric methods in five ways. First, unlike conventional parametric methods, permutation statistical methods are data-dependent methods in that all the information required for analysis is contained within the observed data. Second, permutation methods neither assume nor require random sampling from a defined population, which is essential for parametric methods. Third, permutation methods are distribution-free and do not depend on the usual assumptions associated with conventional parametric methods, such as normality and homogeneity of variance. Fourth, permutation methods provide exact probability values based on the discrete permutation probability distribution, in contrast to parametric methods that provide approximate probability values based on a theoretical approximating distribution. Finally, permutation methods are suitable for small samples, whereas parametric distribution functions often provide very poor fits to the underlying discrete distribution when sample sizes are small.

On the other hand, permutation methods are computationally intensive, oftentimes requiring millions of calculations. A number of calculation efficiencies mitigate the calculation problem, including the recent development of high-speed computing, analyses based on all combinations of the observed data in place of all permutations of the data, the use of mathematical recursion, calculations based on only the variable components of a specified test statistic, and holding constant one treatment array in block designs.

Chapter 4 describes measures of central tendency and variability with which the reader is assumed to be familiar. Emphasized in Chap. 4 is the property of the arithmetic mean as the point about which the sum-of-squared deviations is minimized and the property of the median as the point about which the sum of absolute deviations is minimized. An alternative approach to the mean and median based on paired-squared and paired-absolute differences is introduced.

References

1. Barboza, D., Markoff, J.: Power in numbers: China aims for high-tech primacy. N.Y. Times **161**, D2–D3 (2011)
2. Berkson, J.: Some difficulties of interpretation encountered in the application of the chi-square test. J. Am. Stat. Assoc. **33**, 526–536 (1938)
3. Berry, K.J., Johnston, J.E., Mielke, P.W.: A Chronicle of Permutation Statistical Methods: 1920–2000 and Beyond. Springer, Cham (2014)
4. Biondini, M.E., Mielke, P.W., Berry, K.J.: Data-dependent permutation techniques for the analysis of ecological data. Vegetatio **75**, 161–168 (1988) [The name of the journal was changed to *Plant Ecology* in 1997]
5. Box, G.E.P., Andersen, S.L.: Permutation theory in the derivation of robust criteria and the study of departures from assumption (with discussion). J. R. Stat. Soc. B Methodol. **17**, 1–34 (1955)
6. Bross, I.D.J.: Is there an increased risk? Fed. Proc. **13**, 815–819 (1954)
7. Conover, E.: Quantum computers get real. Sci. News Mag. **191**, 28–33 (2017)
8. Curran-Everett, D.: Explorations in statistics: standard deviations and standard errors. Adv. Physiol. Educ. **32**, 203–208 (2008)
9. Eden, T., Yates, F.: On the validity of Fisher's z test when applied to an actual example of non-normal data. J. Agric. Sci. **23**, 6–17 (1933)
10. Edgington, E.S.: Statistical inference and nonrandom samples. Psychol. Bull. **66**, 485–487 (1966)
11. Feinstein, A.R.: Clinical Biostatistics XXIII: the role of randomization in sampling, testing, allocation, and credulous idolatry (Part 2). Clin. Pharmacol. Ther. **14**, 898–915 (1973)
12. Fisher, R.A.: Statistical Methods for Research Workers. Oliver and Boyd, Edinburgh (1925)
13. Fisher, R.A.: The logic of inductive inference (with discussion). J. R. Stat. Soc. **98**, 39–82 (1935)
14. Geary, R.C.: Some properties of correlation and regression in a limited universe. Metron **7**, 83–119 (1927)
15. Geary, R.C.: Testing for normality. Biometrika **34**, 209–242 (1947)
16. Haber, M.: Comments on "The test of homogeneity for 2×2 contingency tables: a review of and some personal opinions on the controversy" by G. Camilli. Psychol. Bull. **108**, 146–149 (1990)
17. Hays, W.L.: Statistics. Holt, Rinehart and Winston, New York (1988)
18. Hotelling, H., Pabst, M.R.: Rank correlation and tests of significance involving no assumption of normality. Ann. Math. Stat. **7**, 29–43 (1936)
19. Hubbard, R.: Alphabet soup: blurring the distinctions between p's and α's in psychological research. Theor. Psychol. **14**, 295–327 (2004)
20. Hunter, M.A., May, R.B.: Some myths concerning parametric and nonparametric tests. Can. Psychol. **34**, 384–389 (1993)
21. Irwin, J.O.: Tests of significance for differences between percentages based on small numbers. Metron **12**, 83–94 (1935)
22. Johnston, J.E., Berry, K.J., Mielke, P.W.: Permutation tests: precision in estimating probability values. Percept. Motor Skill. **105**, 915–920 (2007)
23. Kempthorne, O.: Why randomize? J. Stat. Plan. Infer. **1**, 1–25 (1977)
24. Kennedy, P.E.: Randomization tests in econometrics. J. Bus. Econ. Stat. **13**, 85–94 (1995)
25. Lachin, J.M.: Statistical properties of randomization in clinical trials. Control. Clin. Trials **9**, 289–311 (1988)
26. Ludbrook, J.: Advantages of permutation (randomization) tests in clinical and experimental pharmacology and physiology. Clin. Exp. Pharmacol. Physiol. **21**, 673–686 (1994)
27. Ludbrook, J.: Issues in biomedical statistics: comparing means by computer-intensive tests. Aust. NZ J. Surg. **65**, 812–819 (1995)
28. Lyons, D.: In race for fastest computer, China outpaces U.S. Newsweek **158**, 57–59 (2011)

29. Matthews, R.: Beautiful, but dangerous. Significance **13**, 30–31 (2016)
30. May, R.B., Hunter, M.A.: Some advantages of permutation tests. Can. Psychol. **34**, 401–407 (1993)
31. Micceri, T.: The unicorn, the normal curve, and other improbable creatures. Psychol. Bull. **105**, 156–166 (1989)
32. Mielke, P.W., Berry, K.J.: Fisher's exact probability test for cross-classification tables. Educ. Psychol. Meas. **52**, 97–101 (1992)
33. Mielke, P.W., Berry, K.J.: Data-dependent analyses in psychological research. Psychol. Rep. **91**, 1225–1234 (2002)
34. Neyman, J., Pearson, E.S.: On the use and interpretation of certain test criteria for purposes of statistical inference: part I. Biometrika **20A**, 175–240 (1928)
35. Neyman, J., Pearson, E.S.: On the use and interpretation of certain test criteria for purposes of statistical inference: part II. Biometrika **20A**, 263–294 (1928)
36. Pitman, E.J.G.: Significance tests which may be applied to samples from any populations. Suppl. J. R. Stat. Soc. **4**, 119–130 (1937)
37. Pitman, E.J.G.: Significance tests which may be applied to samples from any populations: II. The correlation coefficient test. Suppl. J. R. Stat. Soc. **4**, 225–232 (1937)
38. Pitman, E.J.G.: Significance tests which may be applied to samples from any populations: III. The analysis of variance test. Biometrika **29**, 322–335 (1938)
39. Siegfried, T.: Birth of the qubit. Sci. News Mag. **191**, 34–37 (2017)
40. Yates, F.: Contingency tables involving small numbers and the χ^2 test. Suppl. J. R. Stat. Soc. **1**, 217–235 (1934)