JOURNAL OF BIOPHARMACEUTICAL STATISTICS
Vol. 13, No. 4, pp. 735–751, 2003

Sample Size Determination Based on Rank Tests in Clinical Trials

Hansheng Wang,[1] Bin Chen,[2] and Shein-Chung Chow[3,*]

[1]Guanghua School of Management, Peking University, Beijing, P.R. China
[2]Department of Statistics, University of Wisconsin-Madison, Wisconsin, USA
[3]Millennium Pharmaceuticals, Inc., Cambridge, Massachusetts, USA

ABSTRACT

The problem of sample size determination based on three commonly used nonparametric rank based tests, namely, one-sample Wilcoxon's rank sum test, two-sample's Wilcoxon's rank sum test, and the rank-based test for independence is studied. Explicit formulas for variabilities of the test statistics under the alternative hypotheses are derived. Consequently, close forms of power functions of these test statistics are obtained for sample size determination utilizing the concept of higher order polynominal equations. Simulation studies were performed to evaluate the finite samples performance of the derived sample size formulas. The results indicates that the derived methods work well with moderate sample size.

Key Words: Sample size; Nonparametrics; Higher order polynominal equation.

*Correspondence: Shein-Chung Chow, Millennium Pharmaceuticals, Inc., Cambridge, MA 02139, USA; E-mail: shein-chung.chow@mpi.com.

DOI: 10.1081/BIP-120024206
1054-3406 (Print); 1520-5711 (Online)

www.dekker.com

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

1. INTRODUCTION

In clinical research, sample size calculation/justification plays an important role for the validity and success of a clinical trial. The objective of sample size calculation is to estimate the minimum sample size needed for achieving a desired power at a given level of significance. In practice, if a study treatment is truly different from a control, such a statistical difference can always be detected at any significance level if the sample size is sufficiently large. If the sample size is too small, the intended trial may not have sufficient power to detect such a difference. As a result, the common practice for sample size calculation is to select a minimum sample size that can achieve a desired power (e.g., 80%) at a given level of significance (e.g., 5%). On the other hand, the objective of the sample size justification is multifold. First, it is to evaluate how much power the intended trial can achieve for a selected sample size. Second, it is to determine what difference can be detected with the selected sample size for a given desired power. For good clinical practice, it is suggested that sample size calculation/justification should be included in the study protocol before the conduct of a clinical trial, see Chow and Liu (1998) and ICH (1996).

In Clinical research, clinical trials are usually conducted for evaluation of the efficacy and safety of a test drug as compared to a placebo control or an active control agent (e.g., a standard therapy) in terms of mean responses of some primary study endpoints. Under normality assumption with constant variability, the standard analysis of variance (ANOVA) is usually performed to evaluate the treatment effect. In practice, however, it is not uncommon to encounter the situations where the normality assumption is not met even after some data transformation (e.g., log-transformation). In this case, it is recommended that various rank-based nonparametric tests be used for assessment of treatment effect. As compared to the ANOVA, the rank-based tests are usually asymptotically correct with minimum assumptions of the distribution.

In clinical research, however, sample size calculation based on rank-based tests are not well studied in the literature. One of the difficulties for rank-based nonparametric tests is that under the alternative hypothesis, the variability of the test statistic is difficult to derive. The other difficulty is that higher order nonlinear equations are usually involved when solving the required sample size. In this article, we derive formulas for sample size calculation based on the three most commonly used rank-based tests, namely, one-sample rank-sum test, two-sample rank-sum test, and test for independence. We first derive the variabilities of these test statistics under the alternative hypothesis. Then, we provide explicit formulas for sample size calculation. The validity of the derived formulas is confirmed with simulation studies.

The remainder of this article is organized as follows. In the next section, the one-sample rank-sum test for testing location parameter is discussed. Section 3 introduces the two-sample rank-sum test for testing location parameter. In Sec. 4, the test for independence is explored. Also included in each section are simulation results and real examples for illustration of the proposed methods. Details of the proofs are given in the Appendix.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

2. ONE-SAMPLE LOCATION PROBLEM

2.1. Model and Power Analysis

In clinical research, it is often of interest to evaluate whether there is a difference before and after treatment. Thus, our primary interest is to determine whether a shift in location has occurred after the application of treatment. Let x_i and y_i, $i=1,\ldots,n$ be the observations obtained from the ith subject before and after the application of treatment, respectively. Let $z_i = y_i - x_i$, $i=1,\ldots,n$. Consider the following model

$$z_i = \theta + e_i, \quad i = 1,\ldots,n$$

where θ is the unknown location parameter (or treatment effect) of interest and the e_is are random errors in observing z_i. It is assumed that each e_i comes from a continuous population (not necessarily the same one) that is symmetric about zero and the e_is are mutually independent. The hypotheses concerning the location parameter of interest are given by

$$H_0 : \theta = 0 \quad \text{vs.} \quad H_a : \theta \neq 0$$

To test the above hypotheses, Wilcoxon's signed rank test is commonly employed. Consider the absolute differences $|z_i|$, $i=1,\ldots,n$. Let R_i be the rank of $|z_i|$ in the joint ranking from least to greatest. Define

$$\psi_i = \begin{cases} 1 & \text{if } z_i > 0 \\ 0 & \text{if } z_i < 0 \end{cases}$$

where $i=1,\ldots,n$. Then, the statistic

$$T^+ = \sum_{i=1}^{n} R_i \psi_i$$

is the sum of the positive signed ranks. It can be easily verified that under the null hypothesis, the statistic

$$T^* = \frac{T^+ - E(T^+)}{\sqrt{\text{var}(T^+)}} = \frac{T^+ - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}} \tag{1}$$

has an asymptotic standard normal distribution. Therefore, when n is large enough, we may reject the null hypothesis at the α asymptotic level of significance for large n if

$$|T^*| \geq z_{\alpha/2}$$

To determine the sample size, we need to evaluate the mean and variance of T^+ under a given alternative hypothesis. By writing T^+ as a sum of index functions,

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

the mean and variance of T^+ can be obtained as follows

$$\begin{aligned} E(T^+) &= np_1 + n(n-1)p_2 \\ \text{var}(T^+) &= np_1(1-p_1) + n(n-1)(p_1^2 - 4p_1p_2 + 3p_2 - 2p_2^2) \\ &\quad + n(n-1)(n-2)(p_3 + 4p_4 - 4p_2^2) \end{aligned}$$

where

$$\begin{aligned} p_1 &= P(z_1 > 0) \\ p_2 &= P(z_1 \geq |z_2|) \\ p_3 &= P(z_1 \geq |z_2|, z_1 \geq |z_3|) \\ p_4 &= P(z_1 \geq z_2 \geq |z_3|) \end{aligned}$$

Note that the derivation of $\text{Var}(T^+)$ is given in Theorem 1 of the Appendix. The p_is can be estimated by

$$\begin{aligned} \hat{p}_1 &= \frac{1}{n}\sum_{i=1}^{n} I\{z_I > 0\} \\ \hat{p}_2 &= \frac{1}{n(n-1)}\sum_{i\neq j} I\{z_i \geq |z_j|\} \\ \hat{p}_3 &= \frac{1}{n(n-1)(n-2)}\sum_{i\neq j\neq k} I\{z_i \geq |z_j|, z_i \geq |z_k|\} \\ \hat{p}_4 &= \frac{1}{n(n-1)(n-2)}\sum_{i\neq j\neq k} I\{z_i \geq z_j \geq |z_k|\} \end{aligned}$$

Denote $\sigma_+^2 = \text{var}(T^+)$. Then, under the alternative hypothesis, T^+ can be approximated by a normal random variable with mean $E(T^+) = np_1 + n(n-1)\,p_2$ and variance σ_+^2. Note that when the alternative hypothesis is true, $E(T^+) \neq \frac{n(n+1)}{4}$. Without loss of generality, we consider the case when $E(T^+) > n(n+1)/4$. The power of the test in Eq. (1) can be approximated by

$$\begin{aligned} 1-\beta &= \left(|T^*| > z_{\alpha/2}\right) \\ &\approx P\left(T^* > z_{\alpha/2}\right) \\ &= P\left(T^+ > z_{\alpha/2}\sqrt{n(n+1)(2n+1)/24} + \frac{n(n+1)}{4}\right) \\ &= P\left(\frac{T^+ - np_1 - p_2 n(n-1)}{\sigma_+} \right. \\ &\qquad \left. > \frac{z_{\alpha/2}\sqrt{n(n+1)(2n+1)/24} + n(n-1)(1/4 - p_2) + n(1/2 - p_1)}{\sigma^+}\right) \\ &\approx P\left(N(0,1) > \frac{z_{\alpha/2}/\sqrt{12} + \sqrt{n}(1/4 - p_2)}{\sqrt{12(p_3 + 4p_4 - 4p_2^2)}}\right) \end{aligned}$$

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

where the last approximation in the above equation is obtained by ignoring the lower order terms of n. Hence, the sample size required for achieving a desired power of 1-β can be obtained by solving the following equation

$$\frac{z_{\alpha/2}/\sqrt{12} + \sqrt{n}(1/4 - p_2)}{\sqrt{(p_3 + 4p_4 - 4p_2^2)}} = -z_\beta$$

which leads to

$$n = \frac{\left(z_{\alpha/2}/\sqrt{12} + z_\beta\sqrt{p_3 + 4p_4 - 4p_2^2}\right)^2}{(1/4 - p_2)^2} \tag{2}$$

2.2. A Simulation Study

A simulation study was conducted to evaluate the performance of the derived sample size of the formula in Eq. (2). The z_is are generated from normal population with mean θ and variance 1. The p_is are estimated by Monte Carlo method based on a sample of size 10,000. The estimated values of p_is are used to determine the sample size from the formula in Eq. (2). Then, using the calculated sample size, the true power is simulated based on 10,000 simulations. Table 1 summarizes the results from the simulation. As can be seen from Table 1, the sample size needed to achieve the desired power is not too large, and the actual power for the calculated sample size is very close to the nominal power, which indicates that the sample size formula works very well.

2.3. An Example

To illustrate the use of the sample size formula in Eq. (2), we consider an example concerning a clinical study of osteoprosis in postmenopausal women. Suppose a clinical trial is planned to investigate the effect of a test drug on the prevention of the progression to osteoprosis in women with osteopenia. Suppose that a pilot study with five subjects were conducted. The data regarding the bone density before and after the treatment are given in Table 2. It can be estimated that

$$p_2 = 6/20 = 0.30$$

$$p_3 = 4/10 = 0.40$$

$$p_4 = 1/20 = 0.05$$

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

Table 1. Sample size n and actual power for one-sample location with θ and estimated p and nominal power = 0.80, 0.90 (10,000 simulations).

					Nominal power = 0.80		Nominal power = 0.90	
θ	p_1	p_2	p_3	p_4	n	True power	n	True power
0.200	0.579	0.307	0.211	0.060	194	0.776	257	0.878
0.210	0.586	0.309	0.214	0.061	181	0.787	249	0.896
0.220	0.585	0.311	0.213	0.063	173	0.798	234	0.907
0.230	0.591	0.314	0.216	0.062	154	0.792	204	0.886
0.240	0.593	0.315	0.216	0.065	154	0.832	199	0.909
0.250	0.599	0.321	0.221	0.064	124	0.767	183	0.907
0.260	0.604	0.322	0.223	0.067	124	0.804	156	0.881
0.270	0.607	0.323	0.223	0.069	122	0.824	144	0.885
0.280	0.610	0.327	0.227	0.067	104	0.792	133	0.876
0.290	0.614	0.329	0.228	0.070	103	0.813	118	0.852
0.300	0.620	0.335	0.232	0.072	86	0.763	118	0.879
0.310	0.621	0.336	0.233	0.072	85	0.785	121	0.909
0.320	0.625	0.337	0.234	0.074	86	0.811	103	0.886
0.330	0.628	0.340	0.236	0.074	77	0.791	102	0.899
0.340	0.631	0.341	0.237	0.074	76	0.815	101	0.910
0.350	0.640	0.346	0.240	0.076	67	0.780	91	0.895
0.360	0.639	0.346	0.241	0.076	66	0.800	94	0.920
0.370	0.644	0.348	0.242	0.078	65	0.819	83	0.899
0.380	0.646	0.349	0.244	0.080	65	0.838	78	0.899
0.390	0.653	0.356	0.247	0.081	55	0.789	71	0.882
0.400	0.656	0.356	0.248	0.081	54	0.803	71	0.903
0.410	0.658	0.359	0.251	0.083	52	0.804	67	0.898
0.420	0.660	0.362	0.252	0.083	48	0.790	66	0.907
0.430	0.666	0.362	0.252	0.084	49	0.817	63	0.907
0.440	0.671	0.369	0.257	0.086	42	0.774	60	0.905
0.450	0.672	0.368	0.256	0.088	45	0.814	53	0.882
0.460	0.679	0.374	0.262	0.088	38	0.770	51	0.872
0.470	0.679	0.374	0.26	0.088	38	0.784	50	0.888
0.480	0.684	0.375	0.263	0.091	39	0.806	50	0.900
0.490	0.689	0.378	0.265	0.092	37	0.806	46	0.889

Table 2. Data lising of osteopenia pilot study.

Pretreatment	Posttreatment	z
2.87	3.16	0.29
3.03	1.41	−1.62
0.77	3.81	3.04
2.32	3.81	1.49
2.67	1.99	−0.68

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

Hence, the sample size needed to achieve an 80% power for detection of a clinically meaningful improvement is given by

$$\begin{aligned} n &= \frac{\left(z_{\alpha/2}/\sqrt{12} + z_{\beta}\sqrt{p_3 + 4p_4 - 4p_2^2}\right)^2}{(1/4 - p_2)^2} \\ &= \frac{\left(1.96/\sqrt{12} + 0.84\sqrt{0.4 + 4 \times 0.05 - 4 \times 0.3^2}\right)^2}{(0.25 - 0.30)^2} \\ &\approx 383 \end{aligned}$$

Thus, a total of 383 subjects are needed in order to have an 80% power to confirm the observed posttreatment improvement from the pilot study.

3. TWO-SAMPLE LOCATION PROBLEM

3.1. Model and Power Analysis

Let x_i, $i = 1, \ldots, n_1$ and y_j, $j = 1, \ldots, n_2$ be two random samples. One sample (x_i, $i = 1, \ldots, n_1$) from the control population and the other independent sample (y_j, $j = 1, \ldots, n_2$) from the treatment population in a clinical trial. Suppose the primary objective is to investigate whether there is a shift in location (or a treatment effect). Similar to the one-sample location problem, the hypotheses of interest are given by

$$H_0 : \theta = 0 \quad \text{vs.} \quad H_\alpha : \theta \neq 0$$

where θ represents the treatment effect. Consider the following model

$$x_j = e_j, \quad j = 1, \ldots, n_1$$

and

$$y_i = e_{n_1+i} + \theta, \quad i = 1, \ldots, n_2$$

where the es are random variables. It is assumed that each e comes from the same continuous population and the $n_1 + n_2$ es are mutually independent. To test the above hypotheses, Wilcoxon's rank sum test is probably the most commonly used nonparametric test. See, for example, Hollander and Wolfe (1973) and Wilcoxon (1945). To obtain the Wilcoxon′s rank sum test, we first order the $N = n_1 + n_2$ observations from least to greatest and let R_i denote the rank of y_i in this ordering. Let

$$W = \sum_{i=1}^{n_2} R_i,$$

which is the sum of the ranks assigned to the ys. We reject the null hypothesis at the α level of significance if

$$W \geq w(\alpha_2, n_2, n_1)$$

or

$$W \leq n_1(n_2 + n_1 + 1) - w(\alpha_1, n_2, n_1)$$

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

where $\alpha = \alpha_1 + \alpha_2$ and $w(\alpha, n_2, n_1)$ satisfies

$$P(W \geq w(\alpha, n_2, n_1)) = \alpha$$

under the null hypothesis. When both n_1 and n_2 are large, the test statistic

$$W^* = \frac{W - E(W)}{\sqrt{\text{var}(W)}} = \frac{W - \frac{1}{2}n_2(n_2 + n_1 + 1)}{\sqrt{\frac{1}{12}n_1 n_2(n_1 + n_2 + 1)}} \tag{3}$$

is asymptotically distributed as a standard normal distribution when the null hypothesis is true. We therefore reject H_0 at the α asymptotic level of significance if $|W^*| \geq z_{\alpha/2}$.

To obtain the sample size under the alternative hypothesis, write

$$W = \sum_{i=1}^{n_2}\left(\sum_{j=1}^{n_2} I\{y_i \geq y_j\} + \sum_{j=1}^{n_1} I\{y_i \geq x_j\}\right) = \frac{n_2(n_2+1)}{2} + \sum_{i=1}^{n_2}\sum_{j=1}^{n_1} I\{y_i \geq x_j\}$$

Then the mean and variance of W are given by

$$E(W) = \frac{n_1(n_1-1)}{2} + n_1 n_2 p_1$$

and

$$\text{var}(W) = n_1 n_2 p_1(1-p_1) + n_1 n_2(n_1-1)(p_2 - p_1^2) + n_1 n_2(n_2-1)(p_3 - p_1^2)$$

where

$$p_1 = P(y_1 \geq x_1)$$
$$p_2 = P(y_1 \geq x_1 \text{ and } y_1 \geq x_2)$$
$$p_3 = P(y_1 \geq x_1 \text{ and } y_2 \geq x_1)$$

The derivation of the above formulas is given in the appendix.

The above p_is can be estimated by

$$\hat{p}_1 = \frac{1}{n_1 n_2}\sum_{i=1}^{n_2}\sum_{j=1}^{n_1} I\{y_i \geq x_j\}$$

$$\hat{p}_2 = \frac{1}{n_1 n_2(n_1-1)}\sum_{i=1}^{n_2}\sum_{j_1 \neq j_2} I\{y_i \geq x_{j_1} \text{ and } y_i \geq x_{j_2}\}$$

$$\hat{p}_3 = \frac{1}{n_1 n_2(n_2-1)}\sum_{i_1 \neq i_2}\sum_{j=1}^{n_1} I\{y_{i_1} \geq x_i \text{ and } y_{i_2} \geq x_j\}$$

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

When both n_1 and n_2 are large, W can be approximated by a normal random variable with mean

$$\mu w = \frac{n_2(n_2+1)}{2} + n_1 n_2 p_1$$

and variance

$$\sigma_W^2 = n_1 n_2 p_1(1-p_1) + n_1 n_2(n_1-1)\left(p_2 - p_1^2\right) + n_1 n_2(n_2-1)\left(p_3 - p_1^2\right)$$

Note that the derivation of σ_W^2 is given in Theorem 2 of the Appendix.

Under the alternative hypothesis that $\theta \neq 0$, it can be shown that that $p_1 \neq 1/2$. Thus $E(W) \neq \frac{n_1(n_1+1)}{2}$. Without loss of generality, we assume that $p_1 > 1/2$. Under the assumption that $n_1/n_2 \to \kappa$, where $\kappa \varepsilon (0,\ 1)$, the power of the test Eq. (3) can be approximated by

$$\begin{aligned}
1-\beta &= P\left(|W^*| > z_{\alpha/2}\right) \\
&\approx P\left(W^* > z_{\alpha/2}\right) \\
&= P\left(\frac{W - n_2(n_2+1)/2 - n_1 n_2 p_1}{\sigma_W}\right. \\
&\qquad \left. > \frac{z_{\alpha/2}\sqrt{n_1 n_2(n_1+n_2+1)/12} + n_1 n_2(1/2 - p_1)}{\sigma_W}\right). \\
&= P\left(N(0,1) > \frac{z_{\alpha/2}\sqrt{\kappa(1+\kappa)/12} + \sqrt{n_2}\kappa(1/2-p_1)}{\sqrt{\kappa^2\left(p_2 - p_1^2\right) + \kappa\left(p_3 - p_1^2\right)}}\right)
\end{aligned}$$

As a result, the sample size needed to achieve the desired power of $1-\beta$ can be obtained by solving the following equation

$$\frac{z_{\alpha/2}\sqrt{\kappa(1+\kappa)/12} + \sqrt{n_2}\kappa(1/2-p_1)}{\sqrt{\kappa^2\left(p_2 - p_1^2\right) + \kappa\left(p_3 - p_1^2\right)}} = -z_\beta$$

This gives

$$n_1 = \kappa n_2 \quad \text{and} \quad n_2 = \frac{\left(z_{\alpha/2}\sqrt{\kappa(\kappa+1)/12} + z_\beta\sqrt{\kappa^2\left(p_2 - p_1^2\right) + \kappa\left(p_3 - p_1^2\right)}\right)^2}{\kappa^2(1/2-p_1)^2} \tag{4}$$

3.2. A Simulation Study

A simulation study was conducted to evaluate the above formula for sample size calculation. The x_is are generated from normal population with mean 0 and variance 1, y_is are generated from normal population with mean θ and variance 1. The sample size ratio κ is chosen to be 1. The p_is are estimated by a Monte Carlo method based on a sample size of 10,000. The estimated values of p_is are used to determined the sample size from the formula in Eq. (4). Using the calculated sample size, the true power is simulated based on 10,000 simulations. The results are given in Table 3.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

Table 3. Sample size n_2 and actual power for two-sample location with θ and estimated p and nominal power 0.80, 0.90 (10,000 simulations).

				Nominal power = 0.80		Nominal power = 0.90	
θ	p_1	p_2	p_3	n_2	True power	n_2	True power
0.220	0.561	0.397	0.397	352	0.814	471	0.913
0.240	0.570	0.405	0.406	267	0.775	360	0.877
0.260	0.574	0.409	0.413	237	0.789	329	0.905
0.280	0.575	0.412	0.413	231	0.835	290	0.906
0.300	0.585	0.421	0.423	181	0.793	238	0.890
0.320	0.590	0.428	0.430	161	0.800	211	0.891
0.340	0.594	0.432	0.434	147	0.803	174	0.871
0.360	0.602	0.438	0.440	124	0.786	172	0.909
0.380	0.608	0.447	0.449	110	0.779	146	0.882
0.400	0.612	0.452	0.449	103	0.804	131	0.882
0.420	0.618	0.458	0.461	92	0.792	125	0.904
0.440	0.622	0.463	0.465	86	0.799	117	0.901
0.460	0.631	0.473	0.472	75	0.778	101	0.883
0.480	0.630	0.474	0.475	76	0.817	99	0.905
0.500	0.639	0.484	0.483	66	0.801	87	0.896
0.520	0.645	0.489	0.488	60	0.790	81	0.895
0.540	0.645	0.493	0.491	61	0.825	73	0.884
0.560	0.657	0.505	0.503	51	0.783	71	0.900
0.580	0.657	0.505	0.505	51	0.804	66	0.897
0.600	0.666	0.515	0.514	46	0.796	64	0.908
0.620	0.669	0.519	0.520	45	0.817	57	0.890
0.640	0.671	0.523	0.523	43	0.822	55	0.903
0.660	0.678	0.531	0.530	40	0.809	51	0.896
0.680	0.682	0.536	0.535	38	0.813	49	0.903
0.700	0.689	0.543	0.542	35	0.804	47	0.907
0.720	0.694	0.550	0.550	33	0.802	42	0.893
0.740	0.700	0.556	0.557	31	0.794	42	0.904
0.760	0.706	0.565	0.564	29	0.798	39	0.899
0.780	0.710	0.569	0.569	28	0.793	37	0.895
0.800	0.715	0.575	0.574	27	0.802	35	0.894

From Table 3 we see that the sample size needed to achieve the desired power is not too large and the actual power under the calculated sample size is very close to the nominal power, which indicates that the sample size formula in Eq. (4) works very well.

3.3. An Example

To illustrate the use of sample size formula in Eq. (4) derived above, we consider an example concerning a clinical trial for evaluation of the effect of a test drug on

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

Table 4. Data listing of cholesterol pilot study.

Treatment 1(x)	Treatment 2(y)
1.57	3.53
2.31	1.23
0.47	2.15
1.24	2.34
2.78	1.45

cholesterol in patients with coronary heart disease (CHD). Suppose the investigator is interested in comparing two cholesterol lowering agents for treatment of patients with CHD through a parallel design. The primary efficacy parameter is the low density lipoprotein (LDL). The null hypothesis of interest is the one of no treatment difference. Suppose that a two-arm parallel pilot study with five subjects to each arm was conducted. The data regarding the cholesterol pilot study are given in Table 4. It can be estimated that

$$p_1 = 10/25 = 0.40$$
$$p_2 = 10/50 = 0.20$$
$$p_3 = 10/50 = 0.20$$

Hence, the sample size needed in order to achieve an 80% power for detection of a clinically meaningful difference between the treatment groups can be estimated by

$$n = \frac{\left(z_{\alpha/2}/\sqrt{6} + z_{\beta}\sqrt{p_2 + p_3 - 2p_1^2}\right)^2}{(1/2 - p_1)^2}$$
$$= \frac{\left(1.96/\sqrt{6} + 0.84\sqrt{0.20 + 0.20 - 2 \times 0.40^2}\right)^2}{(0.50 - 0.40)^2}$$
$$= 107.69 \approx 108$$

Hence, 108 subjects per arm are needed in order to have an 80% power to confirm the observed difference between the two treatment groups when such a difference truly exists.

4. TEST FOR INDEPENDENCE

4.1. Model and Power Analysis

In many clinical trials, data collected may consist of a random sample from a bivariate population, for example, the baseline value and the posttreatment value. For such data, it is of interest to determine whether there is an association between the two variates involved in the bivariate structure.

In other words, it is of interest to test for independence between the two variates. Let (x_i, y_i), $i=1,\ldots,n$ be the n bivariate observation from the n subjects involved in a clinical trial. It is assumed that (x_i, y_i), $i=1,\ldots,n$ are mutually independent and each (x_i, y_i) comes from the same continuous bivariate population. To obtain a nonparametric test for independence between X and Y, define

$$\tau = 2P((X_1 - X_2)(Y_1 - Y_2) > 0) - 1$$

where τ is the so-called Kendall coefficient. Testing the hypothesis that X and Y are independent is equivalent to test the hypothesis that $H_0: \tau = 0$. Under the null hypothesis, a nonparametric test can be obtained as follows.

First, for $1 \le i < j \le n$, calculate $\zeta(x_i, x_j, y_i, y_j)$, where

$$\zeta(a,b,c,d) = \begin{cases} 1 & \text{if } (a-b)(c-d) > 0 \\ -1 & \text{if } (a-b)(c-d) < 0 \end{cases}$$

Consider

$$K = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \zeta(x_i, x_j, y_i, y_j)$$

We then reject the null hypothesis that $\tau = 0$ at the α level of significance if

$$K \ge k(\alpha_2, n) \quad \text{or} \quad K \le -k(\alpha_1, n)$$

when $k(\alpha, n)$ satisfies that

$$P(K \ge k(\alpha, n)) = \alpha$$

and $\alpha = \alpha_1 + \alpha_2$. Under the null hypothesis, when $n \to \infty$, it can be proved that

$$K^* = \frac{K - E(K)}{\sqrt{\text{var}(K)}} = K\left[\frac{n(n-1)(2n+5)}{18}\right]^{-1/2} \tag{5}$$

is asymptotically distributed as a standard normal distribution. Hence, we would reject the null hypothesis at the α asymptotic level of significance for large samples if $|K^*| \ge z_{\alpha/2}$. It should be noted that when there are ties among the n X observations or among the n Y observations, $\zeta(a, b, c, d)$ should be replaced with

$$\zeta^*(a,b,c,d) = \begin{cases} 1 & \text{if } (a-b)(c-d) > 0 \\ 0 & \text{if } (a-b)(c-d) = 0 \\ -1 & \text{if } (a-b)(c-d) < 0 \end{cases}$$

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

As a result, under H_0, var(K) becomes

$$\begin{aligned}\operatorname{var}(K) = \frac{1}{18}\left[n(n-1)(2n+5) - \sum_{i=1}^{g} t_i(t_i-1)(2t_i+5) - \sum_{j=1}^{h} u_j(u_j-1)(2u_j+5)\right] \\ + \frac{1}{9n(n-1)(n-2)}\left[\sum_{i=1}^{g} t_i(t_i-1)(t_i-2)\right]\left[\sum_{j=1}^{h} u_j(u_j-1)(u_j-2)\right] \\ + \frac{1}{2n(n-1)}\left[\sum_{j=1}^{g} t_i(t_i-1)\right]\left[\sum_{i=1}^{h} u_j(u_j-1)\right]\end{aligned}$$

where g is the number of tied X groups, t_i is the size of the tied X group i, h is the number of tied Y groups, and u_j is the size of the tied Y group j. A formula for sample size calculation can be derived base on test statistic K in Eq. (5). Define

$$\zeta_{i,j} = \zeta(x_i, x_j, y_i, y_j)$$

Then

$$E(K) = \frac{n(n-1)}{2}(2p_1 - 1)$$

and

$$\begin{aligned}\operatorname{var}(K) &= \operatorname{var}\left(\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\zeta_{i,j}\right) \\ &= \frac{n(n-1)}{2}\operatorname{var}(\zeta_{i,j}) + n(n-1)(n-2)\operatorname{cov}(\zeta_{i,j_1}, \zeta_{i,j_2}) \\ &= \frac{n(n-1)}{2}\left[1 - (1-2p_1)^2\right] + n(n-1)(n-2)\left[2p_2 - 1 - (1-2p_1)^2\right]\end{aligned}$$

where

$$\begin{aligned}p_1 &= P((x_1 - x_2)(y_1 - y_2) > 0) \\ p_2 &= P((x_1 - x_2)(y_1 - y_2)(x_1 - x_3)(y_1 - y_3) > 0)\end{aligned}$$

The above p_is can be readily estimated by

$$\begin{aligned}\hat{p}_1 &= \frac{1}{n(n-1)}\sum_{i\neq j} I\{(x_i - x_j)(y_i - y_j) > 0\} \\ \hat{p}_2 &= \frac{1}{n(n-1)(n-2)}\sum_{i\neq j_1\neq j_2} I\{(x_i - x_{j_1})(y_i - y_{j_1})(x_i - x_{j_2})(y_i - y_{j_2}) > 0\}\end{aligned}$$

Under the alternative hypothesis, as $n \to \infty$, it can be shown that K is approximately distributed as a normal random variable with mean

$$\mu_K = \frac{n(n-1)}{2}(2p_1 - 1)$$

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

and variance

$$\sigma_K^2 = \frac{n(n-1)}{2}\left[1-(1-2p_1)^2\right] + n(n-1)(n-2)\left[2p_2 - 1 - (1-2p_1)^2\right]$$

Note that the derivation of σ_K^2 is given in Theorem 3 of the Appendix. Without loss of generality, we assume $p_1 > 1/2$. The power of test in Eq. (5) can be approximated by

$$\begin{aligned}
1-\beta &= P\left(|K^*| > z_{\alpha/2}\right) \\
&\approx P\left(K^* > z_{\alpha/2}\right) \\
&= P\left(\frac{K - n(n-1)(2p_1-1)/2}{\sigma_K}\right. \\
&\qquad \left. > \frac{z_{\alpha/2}\sqrt{n(n-1)(2n+5)/18} - n(n-1)(p_1-1/2)}{\sigma_K}\right) \\
&\approx P\left(N(0,1) > \frac{z_{\alpha/2}/3 - \sqrt{n}(p_1-1/2)}{\sqrt{2p_2 - 1 - (2p_1-1)^2}}\right)
\end{aligned}$$

Hence, the sample size needed in order to achieve a desired power of $1-\beta$ can be obtained by solving the following equation

$$\frac{z_{\alpha/2}/3 - \sqrt{n}(p_1-1/2)}{\sqrt{2p_2 - 1 - (2p_1-1)^2}} = -z_\beta$$

This yields

$$n = \frac{4\left(z_{\alpha/2}/3 + z_\beta\sqrt{2p_2 - 1 - (2p_1-1)^2}\right)^2}{(2p_1-1)^2} \tag{6}$$

4.2. A Simulation Study

Similarly, a simulation study was conducted to evaluate the performance of the above-derived sample size formula. The (x_i, y_i)s are generated in the following way: for any given correlation coefficient ρ, let $x_i = u_i$ and $y_i = \frac{\rho}{\sqrt{1-\rho^2}}u_i + v_i$, where u_i and v_i are random samples generated from the standard normal distribution. The p_is are estimated by Monte Carlo method based on a sample of size 10,000. The estimated values of p_is are used to determine the sample size from the formula in Eq. (6). Then, using the calculated sample size, the true power is simulated based on 10,000 simulations. Table 5 summarizes the results. From Table 5 we see that the sample size needed to achieve the desired power is not too large and the actual power under the calculated sample size is very close to the nominal power, which indicates that the sample size formula in Eq. (6) works very well.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

Table 5. Sample size n and actual power for independence with correlation coefficient ρ and estimated p and nominal power = 0.80, 0.90 (10,000 simulations).

			Nominal power = 0.80		Nominal power = 0.90	
ρ	p_1	p_2	n	True power	n	True power
0.200	0.563	0.561	218	0.809	290	0.905
0.210	0.567	0.563	193	0.799	265	0.908
0.220	0.570	0.563	178	0.806	237	0.900
0.230	0.574	0.563	156	0.789	233	0.925
0.240	0.579	0.564	137	0.766	199	0.903
0.250	0.581	0.562	129	0.786	181	0.902
0.260	0.581	0.565	132	0.815	170	0.909
0.270	0.585	0.568	119	0.802	143	0.876
0.280	0.591	0.569	105	0.791	141	0.892
0.290	0.595	0.568	94	0.774	137	0.911
0.300	0.599	0.570	87	0.766	116	0.877
0.310	0.599	0.571	87	0.796	117	0.905
0.320	0.603	0.569	80	0.797	110	0.899
0.330	0.608	0.572	72	0.775	98	0.884
0.340	0.612	0.574	68	0.771	93	0.886
0.350	0.615	0.578	65	0.777	86	0.882
0.360	0.617	0.577	62	0.781	76	0.863
0.370	0.620	0.578	59	0.792	72	0.864
0.380	0.624	0.580	55	0.786	73	0.896
0.390	0.628	0.583	53	0.784	66	0.875
0.400	0.629	0.580	50	0.783	65	0.881

4.3. An Example

Suppose x and y are two primary responses in a clinical trial. Also, suppose in a pilot study, it is observed that a larger x value resulted in a larger value of y. Thus, it is of interest to conduct a clinical trial to confirm that such an association between two primary responses, x and y, truly exists. Data from the pilot study is given in Table 6.

Table 6. Data listing for testing independence.

x	y
1.42	0.65
0.59	1.58
0.40	0.68
0.27	0.14
0.53	0.59

It is estimated that

$$p_1 = 7/10 = 0.70$$
$$p_2 = 14/15 = 0.93$$

Hence, the sample size required for achieving an 80% power can be obtained as follows

$$n = \frac{4\left(z_{\alpha/2}/3 + z_\beta\sqrt{2p_2 - 1 - (2p_1 - 1)^2}\right)^2}{(2p_1 - 1)^2}$$
$$= \frac{4\left(1.96/3 + 0.84\sqrt{2 \times 14/15 - 1 - (1.40 - 1)^2}\right)^2}{(1.40 - 1)^2}$$
$$= 46.20 \approx 47$$

Thus, 47 subjects per arm are needed in order to achieve an 80% power to confirm the association between x and y observed from the pilot study.

APPENDIX

Theorem 1. *Under the assumptions as described in Sec. 2 for the one-sample location problem, the variance of T^+ is given by*

$$\text{var}(T_+) = np_1(1 - p_1) + n(n-1)(p_1^2 - 4p_1p_2 + 3p_2 - 2p_2^2),$$
$$+ n(n-1)(n-2)(p_3 + 4p_4 - 4p_2^2)$$

where p_is are given in Sec. 2.

Proof.

$$\text{var}(T^+) = n\,\text{var}(I\{z_i > 0\})$$
$$+ \frac{n(n-1)}{2}\text{var}(I\{|z_i| \geq |z_j|, z_i > 0\} + I\{|z_j| \geq |z_i|, z_j > 0\})$$
$$+ 2n(n-1)\text{cov}(I\{z_i \geq 0\}, I\{|z_i| \geq |z_j|, z_i > 0\} + I\{|z_j| \geq |z_i|, z_j > 0\})$$
$$+ \frac{n(n-1)}{2}2(n-2)\text{cov}(I\{|z_i| \geq |z_{j1}|, z_i > 0\} + I\{|z_{j1}| \geq |z_i|, z_{j1} > 0\},$$

$$I\{|z_i| \geq |z_{j2}|, z_i > 0\} + I\{|z_{j2}| \geq |z_i|, z_{j2} > 0\})$$
$$= np_1(1 - p_1) + n(n-1)(p_1^2 - 4p_1p_2 + 3p_2 - 2p_2^2)$$
$$+ n(n-1)(n-2)(p_3 + 4p_4 - 4p_2^2)$$

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

Theorem 2. *Under the assumptions as described in Sec.* 3 *for two-sample location problem, the variance of W is given by*

$$\mathrm{var}(W) = n_1 n_2 p_1(1-p_1) = n_1 n_2(n_1-1)(p_2-p_1^2) + n_1 n_2(n_2-1)(p_3-p_1^2)$$

where p_is are given in Sec. 3.

Proof.

$$\begin{aligned}
\mathrm{var}(W) &= \mathrm{var}\left(\frac{n_2(n_2+1)}{2} + \sum_{i=1}^{n_2}\sum_{j=1}^{n_1} I\{y_i \geq x_j\}\right) \\
&= \mathrm{var}\left(\sum_{i=1}^{n_2}\sum_{j=1}^{n_1} I\{y_i \geq x_j\}\right) \\
&= n_1 n_2 \mathrm{var}(I\{y_i \geq x_j\}) + n_1 n_2(n_1-1)\mathrm{cov}(I\{y_i \geq x_{j1}\}, I\{y_i \geq x_{j2}\}) \\
&\quad + n_1 n_2(n_2-1)\mathrm{cov}(I\{y_{i1} \geq x_j\}, I\{y_{i2} \geq x_j\}) \\
&= n_1 n_2 p_1(1-p_1) + n_1 n_2(n_1-1)(p_2-p_1^2) + n_1 n_2(n_2-1)(p_3-p_1^2)
\end{aligned}$$

Theorem 3. *Under the assumptions as described in Sec.* 4 *for testing independence between the two varaites, the variance of K is given by*

$$\mathrm{var}(K) = \frac{n(n-1)}{2}[1-(1-2p_1)^2] + n(n-1)(n-2)[2p_2 - 1 - (1-2p_1)^2]$$

where p_is are given in Sec. 4.

Proof.

$$\begin{aligned}
\mathrm{var}(K) &= \mathrm{var}\left[\sum_{i=1}^{n-1}\sum_{j=1+1}^{n} \zeta_{i,j}\right] \\
&= \frac{n(n-1)}{2}\mathrm{var}(\zeta_{i,j}) + n(n-1)(n-2)\mathrm{cov}(\zeta_{i,j1}, \zeta_{i,j2}) \\
&= \frac{n(n-1)}{2}[1-(1-2p_1)^2] + n(n-1)(n-2)[2p_2 - 1 - (1-2p_1)^2]
\end{aligned}$$

REFERENCES

Chow, S. C., Liu, J. P. (1998). *Design and Analysis of Clinical Trials*. New York: John Wiley & Sons.

Hollander, M., Wolfe, D. A. (1973). *Nonparametric Statistical Methods*. New York: John Wiley & Sons.

ICH. (1996). *Harmonised Tripartite Guideline: Guideline For Good Clinical Practice*.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics* 1: 80–83.

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016