STATISTICA, anno LXII, n. 4, 2002

AN EXTENSION OF THE ORDINARY PERMUTATION SOLUTION OF THE TWO-SAMPLE LOCATION PROBLEM

Marco Marozzi

1. INTRODUCTION

We are interested in comparing locations of two populations; more precisely, we are interested in testing $H_0 : {}_1X \overset{d}{=} {}_2X$ against the stochastic dominance (one-sided) alternative $H_1 : {}_1X \overset{d}{>} {}_2X$, where X is a continuous random variable. We assume that the distribution functions of ${}_1X$ and ${}_2X$ may differ only in their location parameters δ_1 and δ_2. Thus we can specify the hypotheses also as

$$H_0 : \vartheta = 0 \text{ versus } H_1 : \vartheta > 0 \quad (1)$$

where $\vartheta = \delta_1 - \delta_2$.

2. THE MOST FREQUENTLY USED PARAMETRIC TEST AND THE ORDINARY PERMUTATION SOLUTION

Let ${}_1X_i$ i=1,…,n_1 (${}_2X_i$ i=1,…,n_2) be a random sample taken from the first (second) population. The most frequently used method for testing the hypotheses (1) is the Student's *t* test, which is based on this statistic

$$T = \frac{{}_1\overline{X} - {}_2\overline{X}}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\left(\frac{\sum_{i=1}^{n_1}\left({}_1X_i - {}_1\overline{X}\right)^2 + \sum_{i=1}^{n_2}\left({}_2X_i - {}_2\overline{X}\right)^2}{n-2}\right)}},$$

where ${}_1\overline{X}$ (${}_2\overline{X}$) denotes the mean of the first (second) sample and n=n_1+n_2. This is the uniformly most powerful similar test in testing (1) when the populations are normal with a common variance and T will follow a Student's t distribution with n-2 degrees of freedom if the null hypothesis is true (Lehmann, 1986). The p-value for the Student's test can be computed from the t with n-2 degrees of freedom cumulative distribution. Under assumptions of normality and homoschedasticity, this test is exact, unbiased and consistent as well; but when the underlying population distributions are not normal, these properties are no longer satisfied. It is worth noting that the practitioner who performs the test rarely knows the distribution underlying the data.

A procedure that is valid for any underlying distribution whatsoever, is the ordinary permutation test for comparing two locations. This test is based on

$$PT = \sum_{i=1}^{n_1} {}_1X_i \, .$$

To test $H_0 : \vartheta = 0$ against $H_1 : \vartheta > 0$, the observed test statistic value ${}_0PT$ is compared to the values of the test statistic in all permutations of the combined sample $\mathrm{x} = \left({}_1X_i, i = 1,\ldots,n_1; {}_2X_i, i = 1,\ldots,n_2\right)$. Since PT is symmetric with respect to the order of ${}_1X_i$ and ${}_1X$ and ${}_2X$ are assumed to be continuous random variables, the n! permutations of x give rise to $\binom{n}{n_1}$ almost surely distinct values of PT. Under H_0 each of $\binom{n}{n_1}$ values of PT has a probability of $\binom{n_1+n_2}{n_1}^{-1}$.

Since the permutation distribution of PT is impractical to enumerate (except for small sample sizes), it is generally approximated by taking a random sample of size B from the set of all permutations. Let ${}_bPT$ be the value of PT in the b-th permutation, b=1,…,B. The p-value L_{PT} of the ordinary permutation test is estimated by

$$\hat{L}_{PT} = \frac{\sum_{b=1}^{B} \mathrm{I}\left({}_bPT \geq {}_0PT\right)}{B},$$

where $\mathrm{I}(.)$ stands for the indicator function: $\mathrm{I}\left({}_bPT \geq {}_0PT\right) = 1$ when ${}_bPT \geq {}_0PT$ and $\mathrm{I}\left({}_bPT \geq {}_0PT\right) = 0$ otherwise. It is worth observing that according to the well-known Glivenko-Cantelli theorem, as B tends to infinity $\hat{L}_{PT}$ converges almost surely to L_{PT}, thus $\hat{L}_{PT}$ is a strong-consistent estimator of the p-value. Let $0 < \alpha < 1$ be the nominal significance level, if $\hat{L}_{PT} \leq \alpha$ then H_0 is rejected, otherwise is accepted.

3. THE PROPOSED METHOD

We selected the permutation approach to develop a new procedure for tackling the two-sample problem because of the numerous advantages of permutation tests with respect to parametric tests. See Marozzi (2001) for a detailed discussion on permutation testing. Here it suffices to stress that a sufficient condition for a permutation test to be exact is the exchangeability of the observations under H_0. Moreover, there exist problems which can be treated only within the permutation framework. Regarding those problems that have already been solved by using a parametric method, it is very often possible to obtain good results by using the permutation version of the proposed parametric method. Generally, the performance of this permutation version is close to that of the parametric method when the assumptions behind the latter are met; otherwise its performance could be even better (see Good (2000) and the references therein).

The theory which has been used to develop an extension of the ordinary permutation solution of the two-sample problem is that of the Nonparametric Combination of Dependent Permutation Tests (NPC), due principally to Pesarin (2001). This theory allows us to take advantage of grouping the sample units into two groups. Our purpose is that of taking two aspects of each sample unit into account: the first refers to its observed value; the second to the fact that this value is "large" or "small". For this reason we have decided to group the sample observations into these two groups: that of the observations which are greater than the median $\tilde{M}$ of the combined sample and that of the observations which are less than or equal to $\tilde{M}$. Let ${}_1Y, {}_2Y, \ldots, {}_nY$ denote the order statistics of the combined sample. For n odd $\tilde{M} = {}_{(n+1)/2}Y$. Since we desire a unique value for $\tilde{M}$, if n is even we use $\tilde{M} = \frac{1}{2}\left({}_{n/2}Y + {}_{n/2+1}Y\right)$.

It follows that the hypotheses under study, $H_0 : {}_1X \overset{d}{=} {}_2X$ against $H_1 : {}_1X \overset{d}{>} {}_2X$, may be broken down as

$$H_0 = H_{0s} \cap H_{0g} \text{ against } H_1 = H_{1s} \cup H_{1g}$$

where $H_{0j}: {}_1X_j \overset{d}{=} {}_2X_j$ and $H_{1j}: {}_1X_j \overset{d}{>} {}_2X_j$ with $j \in \{s, g\}$, and where

$$\begin{cases} {}_1X_s = {}_1X \mid {}_1X \le \tilde{M} \\ {}_1X_g = {}_1X \mid {}_1X > \tilde{M} \end{cases} \text{ and } \begin{cases} {}_2X_s = {}_2X \mid {}_2X \le \tilde{M} \\ {}_2X_g = {}_2X \mid {}_2X > \tilde{M} \end{cases}.$$

The testing procedure is carried out in two successive steps. In the first step we test the sub-hypothesis H_{0s} against H_{1s} and H_{0g} against H_{1g}, while in the second step we test the global system of hypotheses by using NPC theory. The partial test statistics are defined as

$$T_s = \sum_{i=1}^{n_{1s}} {}_1X_{si} \text{ and } T_g = \sum_{i=1}^{n_{1g}} {}_1X_{gi} ,$$

where the symbols used have a clear meaning. With respect to $\hat{L}_{PT}$, we have added 1/2 and 1 respectively to the numerator and denominator of the fraction for a mere computational reason. This does not substantially modify the permutation behavior of the global test we are going to define. See Pesarin (2001) for more details on this and other aspects of NPC theory. It should be pointed out that the permutation model is the same as that of the ordinary solution.

The global test statistic we are going to use is defined as

$$T_{sg} = \sum_{j\in\{s,g\}} \Phi^{-1}\left(1 - L_{T_j}\right),$$

where Φ^{-1} is the inverse of the standard normal cumulative distribution function. The test statistic for testing H_0 against H_1 has been obtained through the nonparametric combination of the partial p-values by using the Liptak combining function. Note that the partial test statistics are permutationally equivalent to their p-values.

The observed value of T_{sg}

$${}_0T_{sg} = \sum_{j\in\{s,g\}} \Phi^{-1}\left(1 - L_{T_j}\left({}_0T_j\right)\right)$$

is estimated as ${}_0\hat{T}_{sg} = \sum_{j\in\{s,g\}} \Phi^{-1}\left(1 - \hat{L}_{T_j}\left({}_0T_j\right)\right)$. The distribution of T_{sg} is simulated by using the same permutation results by which we obtained the simulated distribution of T_s and T_g. Thus we obtain a vector of B permutation values of T_{sg}: $\left({}_bT_{sg}; b = 1,\ldots,B\right)$, where

$${}_bT_{sg} = \sum_{j\in\{s,g\}} \Phi^{-1}\left(1 - \hat{L}_{T_j}\left({}_bT_j\right)\right).$$

Large values of the observed test statistic are evidence against the null hypothesis. The p-value of the global test is strong-consistently estimated as

$$\hat{L}_{T_{sg}} = \frac{\sum_{b=1}^{B} \mathrm{I}\left({}_bT_{sg} \geq {}_0\hat{T}_{sg}\right)}{B}.$$

We reject H_0 at significance level $0<\alpha<1$ if $\hat{L}_{T_{sg}} \leq \alpha$.

The rationale for the method just described is based on the possibility of drawing an inference not only about the global null hypothesis but also about the partial null hypotheses, and then obtaining more inferential information from the data. Imagine that one has rejected H_0; using this method one can judge if there is a different influence on the rejection of H_0 from observations belonging to the two groups (that of "large" observations and that of "small" ones). If, for example, H_{0s} has been rejected but H_{0g} accepted, it may be of some interest to investigate the reasons for this result. This investigation may suggest research about covariates to be introduced in the analysis in order to improve the study.

4. A COMPARATIVE SIMULATION STUDY

To evaluate the performance of the test described in the previous section, Monte Carlo simulations were used to estimate the type-I error rate and the power of T_{sg}, PT and Student's test under normality. We refer to normal sampling because in this case the most used parametric method for testing the system of hypotheses under study is optimal. It is then possible to compare the performance of T_{sg} with that of the optimal solution. Student's test is very often used even when it is not valid (Marozzi, 2001), while the ordinary permutation test for testing (1) is always valid. For these reasons we performed a comparative simulation study under normality of these tests.

The computing programs for performing T_{sg} test and all simulations were coded in R language using the free of charge R 1.2.3 program. Random samples varying in size were generated from two independent normal distributed populations (with variance equal to one) by means of the `rnorm` function.

The null hypothesis tested was that the two populations had equal means against the one-sided alternative hypothesis that the first one had a larger mean. We considered three equal sample size configurations: (n_1=n_2=10), (n_1=n_2=20) and (n_1=n_2=40), and four unequal sample size configurations: (n_1=10, n_2=20), (n_1=10, n_2=40), (n_1=20, n_2=10) and (n_1=40, n_2=10). In addition, five different values of the mean difference ϑ were considered. One condition was determined by the null hypothesis ($\vartheta=0$) and the other four used positive values of ϑ that were specified to achieve power near 30%, 46%, 63% and 80% for the ordinary permutation test. Then, for each sample size configuration, one simulation study was used to estimate the type-I error rate and the others to estimate the power. With $\vartheta=0$ ($\vartheta>0$), namely under H_0 (H_1) 4000 (2000) random samples were drawn from each of the two populations and 2000 (1000) permutations of the combined sample were computed.

TABLE 1

Power estimates in percent for the t test, PT and the proposed test at α=5% and for two samples of equal sizes

ϑ	Student's test	*PT*	T_{sg} test
		n_1=10 n_2=10	
0	5.13	5.08	5.25
0.5	28.5	28.4	28.1
0.7	44.0	44.5	42.9
0.9	61.3	61.2	58.7
1.1	76.2	76.2	73.9
		n_1=20 n_2=20	
0	4.68	4.55	4.75
0.35	29.6	29.2	29.7
0.5	46.0	45.6	46.1
0.65	63.4	63.6	62.8
0.8	80.0	79.6	78.3
		n_1=40 n_2=40	
0	5.35	5.33	5.48
0.25	28.7	28.8	29.1
0.35	46.5	45.4	46.1
0.45	64.6	63.7	64.0
0.55	78.8	78.5	78.6

As shown in table 1, for equal sample sizes, the proposed test is practically as powerful as the Student's *t* test and as the ordinary permutation test, even with the smallest sample sizes. The results reported in table 2 and table 3 indicate that the conclusions from the balanced settings apply to the unbalanced ones: the power and the type-I error rate of T_{sg} are not affected by the fact that samples have unequal sizes.

The size estimates show that the proposed test maintained its size (as the other two tests): the type-I error rate of T_{sg} test ranged from 4.75% to 5.48% for the equal size settings and from 4.38% to 5.33% for the unbalanced settings.

The simulation results show that our permutation method for testing $H_0 : {}_1X \overset{d}{=} {}_2X$ against $H_1 : {}_1X \overset{d}{>} {}_2X$ practically entails no loss of power in the context of normal data with respect both to the ordinary permutation solution and the Student's parametric one. This is a very interesting result because, under normal sampling, the Student's test is the most powerful similar test for testing (at a fixed α value) the system of hypotheses at issue, and the ordinary permutation test is almost equivalent to the permutation version of Student's test. In practice, T_{sg} behaves just like the best parametric method and the permutation version of this one. Then, if one adopts T_{sg} one is sure, for what concerns both type-I error rate and power, to test (1) as if one has adopted the Student's method or the ordinary permutation one. The central aspect is that T_{sg}, through processing both the global system of hypotheses and the considered partial systems of hypotheses, allows a deeper inference than that allowed by the other two tests using exactly the same sample information.

TABLE 2

Power estimates in percent for the t test, PT and the proposed test at α=5%
and for two samples of unequal sizes ($n_1 < n_2$)

ϑ	Student's test	*PT*	T_{sg} test
		n_1=10 n_2=20	
0	5.28	5.25	5.33
0.45	32.6	32.6	32.0
0.62	48.0	48.1	47.2
0.78	61.5	61.1	60.9
0.95	76.3	76.5	75.8
		n_1=10 n_2=40	
0	4.58	4.48	4.38
0.4	29.5	30.2	29.6
0.57	49.8	49.2	48.2
0.73	63.7	63.5	61.9
0.9	79.3	78.9	78.5

TABLE 3

Power estimates in percent for the t test, PT and the proposed test at α=5%
and for two samples of unequal sizes ($n_1 > n_2$)

ϑ	Student's test	*PT*	T_{sg} test
		n_1=20 n_2=10	
0	5.08	5.00	5.30
0.45	31.7	31.9	32.1
0.62	46.1	45.9	45.5
0.78	62.9	62.5	60.9
0.95	78.2	77.8	75.5
		n_1=40 n_2=10	
0	5.40	5.38	5.20
0.4	29.4	29.4	29.7
0.57	48.7	48.4	48.7
0.73	64.7	64.2	64.3
0.9	78.8	78.1	76.1

The same simulation results show also that the proposed test, besides being exact, appears to be unbiased and consistent.

5. CONCLUSION

Permutation tests are not generally endorsed very enthusiastically by either theoretical or applied statisticians, but these tests can be very useful in many contexts. The simulation results showed that the proposed exact permutation test is, within the considered context, as powerful as the Student's *t* test and as the ordinary permutation test. In addition, the practitioner who uses the proposed test can draw inferential conclusions that cannot be drawn by using the usual methods.

For these two reasons the practitioner should take into consideration the use of the proposed permutation test. On the one hand, its performance is practically the same as that of the optimal parametric test (even when the sample sizes are small and when they are very different). On the other hand, through processing both the global system of hypotheses and the considered partial systems of hypotheses, it gives more information on the hypotheses under testing than that given by usual tests.

We presented a method which gives more information about the studied null hypothesis than the usual ones. Additional research is needed regarding the ways of using this type of information profitably.

Dipartimento di Scienze statistiche "Paolo Fortunati" MARCO MAROZZI
Università di Bologna

REFERENCES

P. GOOD, (2000), *Permutation tests: a practical guide to resampling methods for testing hypotheses* 2nd ed., Springer-Verlag, New York.

E.L. LEHMANN, (1986), *Testing statistical hypotheses* 2nd ed., John Wiley, New York.

M. MAROZZI, (2001), *Some notes on nonparametric inferences and permutation tests*, "Metron", forthcoming.

F. PESARIN, (2001), *Multivariate permutation tests with applications in biostatistics*, John Wiley, Chichester.

RIASSUNTO

Un'estensione della soluzione ordinaria di permutazione per il two-sample location problem

In questo lavoro viene proposto un metodo permutazionale per il confronto di locazione tra due popolazioni. Per come è stato congegnato, questo test permette di condurre un'inferenza maggiormente informativa di quella ottenibile con gli altri metodi, usando esattamente le stesse informazioni campionarie. Inoltre i risultati di uno studio di simulazione mostrano come nei contesti considerati esso sia de facto caratterizzato dalla stessa potenza della soluzione ottimale parametrica e come la usa ampiezza sia prossima al livello di significatività nominale.

SUMMARY

An extension of the ordinary permutation solution of the two-sample location problem

In this paper, a permutation method for comparing locations of two populations has been proposed. Due to the way it has been devised, this test allows a deeper inference on testing the studied system of hypotheses than that allowed by the other methods using exactly the same sample data. Furthermore, simulation results show that, in the considered contexts, it performs practically as well as the optimal parametric solution in terms of power and it maintains its size close to the nominal significance level.