4.3 Mann-Whitney U test

Overview

The Mann-Whitney U test was described by Wilcoxon (1945) and Mann & Whitney (1947) . Here, interest lies in comparing two groups of independent samples. This is a non-parametric analogue to a classical two-sample (unpaired) *t*-test.

The null hypothesis

Suppose we have independent response values for a quantity of interest measured from each of two groups. We consider these values in the two groups to be representative observations from each of two random variables, Y_1 and Y_2, respectively. The general null hypothesis tested by the Mann-Whitney U test is:

- H_{01}: the distribution of Y_1 is equivalent to that of Y_2.

Thus, Y_1 and Y_2 are *exchangeable* with one another if H_{01} is true. If we want to assume that the distributions of Y_1 and Y_2 have the same scale, shape, etc., and can only differ in the value of their central location, then we can have a more specific null hypothesis:

- H_{02}: the distribution of Y_1 is *stochastically no larger or smaller than* that of $Y_2;$ or (even more strictly)
- H_{03}: the median of Y_1 = the median of Y_2.

The corresponding (two-sided) alternative hypotheses, in each case, would be phrased as:

- H_{A1}: the distributions of Y_1 and Y_2 differ from one another.
- H_{A2}: the distribution of Y_1 is *stochastically either larger or smaller than* that of $Y_2;$ or (even more strictly)
- H_{A3}: the median of Y_1 $\neq$ the median of Y_2,

respectively.

One may also choose to do a **one-tailed test**, e.g., to assert a more specific alternative hypothesis; namely that Y_1 is stochastically larger than Y_2 (or, the median of Y_1 > the median of Y_2).

Description of the test statistic

Let $\left[y_{1},\ldots,y_{n_1} \right] $ be an independent and identically distributed (i.i.d.) sample of n_1 units from Y_1 ('group 1'), and let $\left[y_{(n_1+1)},\ldots,y_{(n_1+n_2)} \right] $ be an i.i.d. sample of n_2 units from Y_2 ('group 2'). The combined set of sampling units from both groups is therefore given by vector $\mathbf{y} = \left[y_{1},\ldots,y_{(n_1+n_2)} \right]. $ Let vector $\mathbf{r} = \left[r_1,\ldots,r_{n_1+n_2} \right]$ be the corresponding ranks of all the sample values in $\mathbf{y}$ from both groups combined, such that the smallest value obtains rank = 1 and the largest value obtains rank = (n_1+n_2). Next, define R_1 and R_2 as the sum of the ranks for group 1 and group 2, respectively; i.e.

$$ R_1 = \sum_{i=1}^{n_1} r_i \hspace{1cm} \text{and} \hspace{1cm} R_2 = \sum_{i=(n_1+1)}^{(n_1+n_2)} r_i $$

then calculate U_1 and U_2 as follows:

$$ U_1 = n_1n_2+\frac{n_1(n_1+1)}{2} - R_1 \hspace{1cm} \text{and} \hspace{1cm} U_2 = n_1n_2+\frac{n_2(n_2+1)}{2} - R_2 $$

Now, for any given dataset, the calculated values of U_1 and U_2 are not independent of one another. More specifically, $U_1+U_2 = n_1n_2$, so either U_1 or U_2 can be used as a suitable test statistic here.

PRIMER uses U_2 as the test-statistic, but will calculate and provide values for both U_1 and U_2 in the output file for this test.

Treatment of ties

If any values of y_i are equal to one another, then they are tied with one another in terms of rank. Then, if we order the data from smallest to largest and generate corresponding integers in order from 1 to (n_1+n_2) , PRIMER will simply *average* the corresponding ordered integers for any tied values to obtain an average and equivalent rank value for them. Thus, for example, suppose we have two groups with $n_1 = n_2 = 3$ having the following values: $\mathbf{y} = \{ 2, 4, 5, 5, 10, 15 \}$, then the ordered integers are $\{ 1, 2, 3, 4, 5, 6 \}$ and the ranks would be $\mathbf{r} = \{ 1, 2, 3.5, 3.5, 5, 6 \}$, because the 3rd and 4th-ranked values are tied.

Importantly, the presence of ties poses no problem for calculating *p*-values in PRIMER, as the permutation algorithm simply proceeds in the usual way, and an exact test is achieved directly. This contrasts with other available software implementations of the Mann-Whitney U test (e.g., 'wilcox.test' in R), which do not compute an exact *p*-value if there are ties.

Calculation of the *P*-value

Assuming only *exchangeability* of the observations across the two groups, we shall generate the distribution of U_2 under a true null hypothesis by permuting the observations freely across the two groups (always retaining the original sample sizes of n_1 and n_2), which permits a direct empirical calculation of an appropriate *p*-value.

If there are no tied values, then the distributions of U_1 and U_2 under a true null hypothesis are: (i) symmetric; and (ii) identical to one another. However, if there are any *ties*, then these two permutation distributions are neither symmetric nor identical. They will, nevertheless, be *mirror* *images* of one another. Specifically, the right-hand tail of U_2 will mirror the left-hand tail of U_1, and the right-hand tail of U_1 will mirror the left-hand tail of U_2. Because of this mirroring, even in the presence of ties, we still only need one or other of U_1 or U_2 to calculate a correct *p*-value under permutation to achieve an exact test.

One-tailed test

First, we shall consider a one-tailed test. Suppose we have the alternative hypothesis:

- H_A: the distribution of Y_1 is stochastically *larger* than that of $Y_2.$

In this case, we would expect that the observed ranks, r_i, associated with group 1 would typically be larger numbers than those associated with group 2, and therefore (given the above equations), that the value of U_2 would tend to be greater than that of U_1.

Having obtained an observed value of the test-statistic, U_2, for a given dataset, we can then obtain a *p*-value empirically using a permutation algorithm. Specifically, under the simple null hypothesis that the two groups are exchangeable, we can randomly re-order (shuffle) all of the values of y_i in the combined vector $\mathbf{y}$ to yield a vector of permuted data, $\mathbf{y}^\pi$ of same length as the original, (n_1+n_2). The concomitantly re-ordered ranks associated with this permuted vector, $\mathbf{r}^\pi$, are then used to calculate a value of the test-statistic under permutation U_2^π. Repeating this permutation procedure a large number of times (e.g., say n_{perm} = 9999) yields a distribution of values of $U_{2,k}^\pi$ $k = 1,\ldots,n_{\text{perm}}$ under a true null hypothesis.

The *p*-value is then calculated as:

$$ P = \frac{ \sum_{k=1}^{n_\text{perm}} (\text{I}(U_{2,k}^\pi \geq U_2) + 1)}{(n_\text{perm} + 1)} $$ with $\text{I}(\textit{expression}) = 1$ if $\textit{expression}$ is true and zero otherwise. The '$+1$' in the numerator and denominator of this fraction is there to acknowledge the inclusion of the observed value as a member of the permutation distribution. The above expression tallies only the values of $U_{2,k}^\pi$ that equal or positively exceed U_2 in the right-hand tail of the permutation distribution.

Note that the arbitrary ordering of the names of the two groups being compared (i.e., as 'group 1' and 'group 2') can simply be swapped if we wish to perform the test with the alternative hypothesis that Y_1 is stochastically *smaller* than Y_2 (focused on the other tail).

Two-tailed test

For an exact two-tailed test that accommodates ties (hence asymmetry in the permutation distributions), we could examine the permutation distributions for both U_1 and U_2. Specifically, for example, if $U_1 < U_2$, we can calculate the two-tailed *p*-value as the sum of the two relevant individual tail probabilities (i.e., the lower tail of U_1^π and the upper-tail of U_2^π), thusly: $$ P = \text{Pr} (U_1^\pi \leq U_1) + \text{Pr} (U_2^\pi \geq U_2) $$

However, taking advantage of the fact that U_1 and U_2 are not independent of one another, and that the permutation distributions of U_1 and U_2 will be precise mirror images of one another (even if some values are tied and they are not therefore symmetric), we can always calculate the correct two-tailed *p*-value using only the permutation distribution of U_2 as follows:

- If $U_1 < U_2$, then $$ P = 2 \times \frac{ \sum_{k=1}^{n_\text{perm}} (\text{I}(U_{2,k}^\pi \geq U_2) + 1)}{(n_\text{perm} + 1)} $$
- If $U_2 < U_1$, then $$ P = 2 \times \frac{ \sum_{k=1}^{n_\text{perm}} (\text{I}(U_{2,k}^\pi \leq U_2) + 1)}{(n_\text{perm} + 1)} $$

Revision #182
Created 19 November 2024 03:03:28 by Marti
Updated 17 November 2025 01:20:28 by Marti