4.5 Kruskal-Wallis test

Overview

The Kruskal-Wallis test was described by Kruskal (1952) and Kruskal & Wallis (1952) . Its purpose is to compare two or more independent groups of samples and it is an extenstion of the Mann-Whitney U test. It operates on ranked values and will indeed yield an equivalent result to the Mann-Whitney U test in the case of two groups. It is a nonparametric statistical test whose classical counterpart is a one-way analysis of variance (ANOVA).

The null hypothesis

Like the Mann-Whitney U test, the Kruskal-Wallis test operates on ranks. Suppose we have a factor 'A' with $i=1,...,a$ distinct levels (groups or populations), and that there are n_i values of a given random response variable Y, sampled from each of these groups; so, y_{ij} indicates the jth sample value ($j = 1,...,n_i$) drawn from the ith group, and there are a total of $N = \sum_{i=1}^{a}n_i$ values being evaluated in the test. The general null hypothesis being tested here is:

- H_0: *There are no differences* in the distribution of values in the underlying populations represented by the groups.

The alternative hypothesis is:

- H_A: *At least two of the groups differ from one another* in the distribution of values in the underlying populations represented by the groups.

The Kruskal-Wallis test generally assumes that: (i) the sampled values are independent of one another, (ii) the sampled values in each group are drawn at random from each of their respective populations, and (iii) the response variable is continuous. We avoid any assumption that the underlying population values for each group are normally distributed. The null hypothesis just asserts that the values from different groups come from the same underlying population distribution, with identical shape and scale, whatever that distribution may be.

In practice, the test also can be applied to random variables that are discrete (so not necessarily continuous) and the ranks of any tied values can be replaced with their average rank value. Our use of a permutation algorithm to perform the test means an exact test (where the type I error of the test is equal to the *a priori* chosen significance level) is achieved. If we restrict the alternative hypothesis to a shift in location only, we may assert the null and alternative hypotheses for the Kruskal-Wallis test as follows:

- H_0: *There are no differences in the median values* of the underlying populations represented by the groups.
- H_A: *At least two groups differ in the median values* of the underlying populations represented by the groups.

Description of the test-statistic

The first step is to rank all of the values in the full set of data, combined, regardless of their group membership. Let the combined set of sampling units from all groups be denoted by a vector $\mathbf{y} = \left[y_{11}, y_{12}, \ldots,y_{ij},\ldots,y_{a,n_a} \right]$, of length N. Then, let vector $\mathbf{r} = \left[r_{11}, r_{12}, \ldots,r_{ij},\ldots,r_{a,n_a} \right]$ be the corresponding ranks of all the sample values in $\mathbf{y},$ such that the smallest value obtains rank = 1 and the largest value obtains rank = N.

Next, let R_i be the sum of the ranks for any group i, i.e.

$$ R_i = \sum_{j=1}^{n_i} r_{ij} $$

The Kruskal-Wallis test statistic is then defined as:

$$ H = \frac { 12 } { N(N+1) } \sum_{i=1}^a \frac { R_i^2 } { n_i } - 3(N+1) $$

Treatment of ties

The distribution of H under a true null hypothesis is affected by the existence of ties. Thus, in the event of ties, average ranks are calculated, and the H test-statistic is adjusted as described by Kruskal & Wallis (1952) and outlined below.

Calculating average ranks

If any values of y_{ij} are equal, then they are tied with one another in terms of their rank. For any tied values, PRIMER calculates an average rank for them, just as described for the Mann-Whitney U test. So, for example, if we have the following set of values: $$ \mathbf{y} = \{9, 12, 12, 14, 14, 14, 30\}, $$

the set of ordered integers for these 7 sample values would look like this: $$ \{1, 2, 3, 4, 5, 6, 7\} $$

and the rank values $\mathbf{r}$ that we would actually use for the analysis (obtained by replacing the ordered integers with their averages, calculated separately for each set of tied values) would look like this: $$ \mathbf{r} = \{1, 2.5, 2.5, 5, 5, 5, 7\} $$

As previously noted, 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. Note that, for cases where there is a small number of unique values of the test statistic under

permutation, although the *p*-value is *accurate* (there is no bias), its *precision* and hence its utility will depend on the number of unique values of the test-statistic that can be computed for a given problem.

Adjustment to the test-statistic

In the event of tied values, let the number of sets of tied values be g. Within each set $\ell = 1, \ldots, g$, there are t_ℓ tied values. For every set ℓ, we also calculate $T_\ell = (t_\ell - 1)t_\ell(t_\ell + 1)$. Then the H test-statistic is adjusted by dividing it by quantity D:

$$ H_{\text{adj}} = H/D $$

where D is defined as $$ D = 1 - \frac{ \sum_\ell^g T_\ell }{ N^3 - N} $$

We note in passing that this adjustment is not necessary in our case, as the *p*-value is calculated using a permutation approach (see the following section), rather than relying on tabled values. However, PRIMER does calculate H_{adj} in the event of ties, essentially to maintain consistency with results that would be obtained using other packages and to remain true to the description of the test-statistic as described by Kruskal & Wallis (1952) .

Calculating a *p*-value

Having obtained an observed value of the test-statistic, H_{obs}, for a given dataset, we can then obtain a *p*-value empirically using a permutation algorithm. Specifically, under the null hypothesis that all groups are exchangeable, we can randomly re-order (shuffle) all of the values of y_{ij} in the combined vector $\mathbf{y}$ to yield a vector of permuted data, $\mathbf{y}^\pi$ of same length (N) as the original. 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 H^π. Repeating this permutation procedure a large number of times (e.g., say n_{perm} = 9999) yields a large number of values of H^π realised under a true null hypothesis. The probability (*p*-value) associated with the null hypothesis is then estimated empirically as the proportion of values of H^π that are equal to or larger than the observed value of the test-statistic, H_obs. Specifically, if we let H_k^π be the value of H^π obtained for the kth permutation ($k = 1, \ldots, n_\text{perm}$), the *p*-value is calculated as the proportion of H_k^π values that equal or exceed H_obs; i.e., $$ p = \frac{ \sum_{k=1}^{n_\text{perm}} (\text{I}(H_k^\pi \geq H_\text{obs}) + 1)}{(n_\text{perm} + 1)} $$

with $\text{I}(\textit{expression}) = 1$ if $\textit{expression}$ is true and zero otherwise. Note that 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 distribution of H^π under a true null hypothesis.

Revision #126
Created 19 November 2024 03:04:45 by Marti
Updated 29 November 2025 16:13:28 by Marti