Krushkal Wallis H test

INSTRUCTOR: SANTOSH CHHATKULI

Function of the test

This test is developed by W.H. Kruskal and W.A. Wallis and it is also known as Kruskal-Wallis H test. It is a generalization of the Wilcoxon rank sum test or Mann-Whitney U test to the case of $K > 2$ samples, where k = no. of independent samples. It is used to test null hypothesis H_0 that k independent samples have same medians or k independent samples have been drawn from the populations with same distribution with respect to median. The test is non-parametric alternative of one-way analysis of variance.

Test Assumptions

1. Samples are drawn randomly and independently from their respective populations.
2. Dependent variable should be measured at the ordinal or continuous level.
3. Independent variable should consist of three or more categorical, independent groups.
4. Distributions in each group (i.e., the distribution of scores for each group of the independent variable) have the same basic shape (which also means the same variability)

Hypothesis

Null hypothesis assumes that the samples are from identical populations and alternative hypothesis assumes that the samples come from different populations.

H_0: $M_1 = M_2 = \ldots = M_k$ (medians of all groups are equal or k distributions are identical)

H_1: At least population median of one group is different from the population median of at least one other group (At least one distribution is different)

Test statistic (small sample case)

Small sample case: ($n_i < 5$, $i = 1, 2, \ldots, k$**)**

If one of the sample has size which is less than 5, we consider small. The procedure for finding test statistics is as follows:

Let n_i be the number of observations in the i^{th} sample, $i = 1, 2, \ldots, k$

Step 1: Combine all data from k samples into a single series. The number observations in combined sample or pooled sample will be $n = n_1 + n_2 + \ldots + n_k$.

Step 2: Arrange all observations in ascending order (or descending order)

Step 3: Assign rank to sorted observations from 1 to n.

Step 4: Sum up the different ranks for each of the different groups.

R_i = Sum of ranks for i^{th} sample or group ($i = 1, 2, \ldots k$)

Small sample case $(n_i < 5, i = 1, \dots, k)$

The test statistic is given by,

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

$$= \frac{12}{n(n+1)} \left\{ \frac{R_1^2}{n_1} + \cdots + \frac{R_k^2}{n_k} \right\} - 3(n+1)$$

Where,

H = Kruskal-Wallis test statistic

n = Total no. of observations in combined samples ($n = n_1 + n_2 + \dots + n_k$)

R_i = Sum of ranks for i^{th} sample

Correction for ties

The corrected H for ties is given by,

$$H_c = \frac{H}{c}$$

Where,

H_C = Corrected value of H

H = Uncorrected value of H

C = Correction for ties = $1 - \frac{\sum_{j=1}^{k}(t_j^3 - t_j)}{n^3 - n}$

Where tj is the number of ties in each tie group.

If there are few ties, then there will be small difference between HC and H.

Test Statistic (Large sample case)

Large sample case: ($n_i \geq 5$, i = 1, 2, …, k**)**

If each sample size is at least 5, then sampling distribution of H statistic is a chi-square distribution with k – 1 degrees of freedom where k stands for number of independent samples.

$$H \sim \chi^2 \ (k - 1)$$

Decision Rule: Small sample case

For small sample case the critical value of H is obtained from the exact distribution table of H statistic. Let $H_\alpha(n_i)$ be the critical value of H from the table of H distribution. We will use following decision rule:

Reject H_0 if cal H $\geq H_\alpha(n_i)$

Decision Rule: Large sample case

For large sample case we compute the critical value of Chi-square for k –1 degrees of freedom with upper tail probability of α. Let $\chi^2_\alpha(\mathrm{k}-1)$ be the critical value of Chi-square distribution, then we will use following decision rule:

Reject H_0 if cal H ≥ $\chi^2_\alpha(k-1)$