**3.1 Basic Concept**

Kruskal-Wallis test, proposed by Kruskal and Wallis, is a nonparametric method for testing whether samples are originated from the same distribution. It extends the Mann-Whitney *U* test to more than two groups. The null hypothesis of the Kruskal-Wallis test is that the mean ranks of the groups are the same. As the nonparametric equivalent one-way ANOVA, Kruskal-Wallis test is called one-way ANOVA on ranks. Unlike the analogous one-way ANOVA, the nonparametric Kruskal-Wallis test does not assume a normal distribution of the underlying data. Thus, Kruskal-Wallis test is more suitable for analysis of data. Because the post sequencing data are often not normally distributed and contain some strong outliers, it is more appropriate to use ranks rather than actual values to avoid the testing being affected by the presence of outliers or by the non-normal distribution of data.

A popular nonparametric test to compare outcomes among more than two independent groups is the Kruskal Wallis test.   The Kruskal Wallis test is used to compare medians among k comparison groups (k > 2) and is sometimes described as an ANOVA with the data replaced by their ranks.   The null and research hypotheses for the Kruskal Wallis nonparametric test are stated as follows:

H0: The k population medians are equal versus

H1: The k population medians are not all equal

The procedure for the test involves pooling the observations from the k samples into one combined sample, keeping track of which sample each observation comes from, and then ranking lowest to highest from 1 to N, where N = n1+n2 + ...+ nk.

**3.2 Assumptions of Kruskal Wallis Test**

* Independence of Observations – Each observation can belong to only one level.
* Sufficient sample sizes
* No assumption of normality.
* The samples are [independent](https://www.quality-control-plan.com/StatGuide/sg_glos.htm#independent) of each other
* Data in each group is randomly selected

**3.3 Steps Used** [**Kruskal–Wallis test**](https://www.sciencedirect.com/topics/mathematics/kruskal-wallis-test)

The [Kruskal–Wallis test](https://www.sciencedirect.com/topics/mathematics/kruskal-wallis-test) is just the [rank-sum test](https://www.sciencedirect.com/topics/mathematics/rank-sum-test) extended to more than two samples. Think of it informally as testing if the distributions have the same median. The chi-square (χ2) approximation requires five or more members per sample.

1. Name the number of samples m (3, 4, …).
2. Name the sizes of the several samples n1, n2, …, nm; n is the grand total.
3. Combine the data, keeping track of the sample from which each datum arose.
4. Rank the data.
5. Add up the ranks of the data from each sample separately.
6. Name the sums T1, T2, …, Tm.
7. Calculate the Kruskal–Wallis H statistic, which is distributed as chi square, by

Obtain the p-value (as if it were α) from Table (χ2 right tail) for m – 1 degrees of freedom (df).

Where,

N = Number of values obtained from every grouped samples,

= Summation of ranks taken from a particular sample

=Number of values from the equivalent sum of rank.

While performing the H-test, the degree of freedom which is written as df is determined by means of the formula:

Where,

***df*** stands for degrees of freedom

***k*** for the number of groups or samples

Then p-value is calculated

If P-Value is less than 5% or greater than 10%, reject null hypothesis.

If p-Value lie between 5% and 10% accept null hypothesis

**3.4 Outline of proposed approach**

Check all assumption are required [Kruskal–Wallis test](https://www.sciencedirect.com/topics/mathematics/kruskal-wallis-test)

All assumption Correct

No

Rank all of the scores the procedure for ranking is as follows the lowest score gets the lowest rank

Yes

If two or more scores are the same then they are "tied". "Tied" scores get the average of the ranks that they w*ould* have obtained

Add up the ranks of the data from each sample separately. Name the sums T1, T2, …, Tm.

Calculate the Kruskal–Wallis H statistic, which is distributed as chi square, by

Find the degree of freedom

We compare our obtained value of *H* to each of the critical values in that row of the table and draw the conclusion

**3.5 Illustrate with example**

This test is appropriate for use under the following circumstances:

(a) We have three or more conditions that we want to compare;

(b) Each condition is performed by a different group of participants; i.e. we have an independent-measures design with three or more conditions.

(c) The data do not meet the requirements for a parametric test. (i.e. use it if the data are not normally distributed; if the variances for the different conditions are markedly different; or if the data are measurements on an ordinal scale).

Does physical exercise alleviate depression?

We find some depressed people and check that they are all equivalently depressed to begin with. Then we allocate each person randomly to one of three groups: no exercise; 20 minutes of physical exercise per day; or 40 minutes of physical exercise per day. At the end of a month, we ask each participant to rate how depressed they now feel, on a scale that runs from 1 ("totally miserable") through to 100 (ecstatically happy").

Table 3.1 Physical exercise per day

|  |  |  |  |
| --- | --- | --- | --- |
| **S No** | **No exercise** | **Physical exercise** **for**  **20 minutes** | **Physical exercise for**  **40 minutes** |
| 1 | 23 | 22 | 59 |
| 2 | 26 | 27 | 66 |
| 3 | 51 | 39 | 38 |
| 4 | 49 | 29 | 49 |
| 5 | 58 | 46 | 56 |
| 6 | 37 | 48 | 60 |
| 7 | 29 | 49 | 56 |
| 8 | 44 | 65 | 62 |

The appropriate test here is the Kruskal-Wallis test. We have three separate groups of participants, each of whom gives us a single score on a rating scale. Ratings are examples of an ordinal scale of measurement, and so the data are not suitable for a parametric test. The Kruskal-Wallis test will tell us if the differences between the groups are so large that they are unlikely to have occurred by chance. Here are the data:

Table 3.2 Physical exercise per day with mean and SD

|  |  |  |  |
| --- | --- | --- | --- |
| **S No** | **No exercise** | **Physical exercise for**  **20 minutes** | **Physical exercise for**  **40 minutes** |
| 1 | 23 | 22 | 59 |
| 2 | 26 | 27 | 66 |
| 3 | 51 | 39 | 38 |
| 4 | 49 | 29 | 49 |
| 5 | 58 | 46 | 56 |
| 6 | 37 | 48 | 60 |
| 7 | 29 | 49 | 56 |
| 8 | 44 | 65 | 62 |
| Mean | 39.63 | 40.63 | 55.57 |
| SD | 12.85 | 14.23 | 8.73 |

Rank all of the scores, ignoring which group they belong to. The procedure for ranking is as follows: the lowest score gets the lowest rank. If two or more scores are the same then they are "tied". "Tied" scores get the average of the ranks that they *would* have obtained, had they not been tied. Here's the scores again, now with their ranks in brackets:

In detail, this is how the ranks are arrived at for these scores.

(a) "22" is the lowest score. This gets a rank of 1.

(b) "23" is the next lowest score. This gets a rank of 2.

(c) "26" is the next lowest score. This gets a rank of 3.

(d) "27" is the next lowest score. This gets a rank of 4.

(e) There are two instances of "29". This is a "tie". They both get the average of the ranks that they would have been allocated, had they been different from each other. So the next two ranks are 5 and 6. The average of 5 and 6 is 11/2 = 5.5. Both instances of "29" therefore get a rank of 5.5.

(f) "37" is the next lowest score. This gets a rank of 7 (because we've just "used up" ranks 5 and 6).

(g) "38" is the next lowest score, and it gets a rank of 8.

(h) "39" is the next lowest score, and it gets a rank of 9.

(i) "44" gets a rank of 10, "46" gets a rank of 11, and "48" gets a rank of 12.

(j) There are three instances of "49", so this is another tie. They each get the average of the next three unused ranks ( (13+14+15) / 3 = 14).

(k) "51" is the next lowest score, and it gets the next "unused" rank, which is 16.

(l) There are two instances of "56", so they get the average of the next two unused ranks ( (17+18) /2 = 17.5).

(m) "58" gets the next unused rank, which is 19.

(n) "59" gets a rank of 20, "60" gets 21, "62" gets 22, "65" gets 23, and 66 gets 24.

Table 3.3 Assigning rank to each group

|  |  |  |  |
| --- | --- | --- | --- |
| **S No** | **No exercise** | **Physical exercise for**  **20 minutes** | **Physical exercise for**  **60 minutes** |
| 1 | 23(2) | 22(1) | 59(20) |
| 2 | 26(3) | 27(4) | 66(24) |
| 3 | 51(16) | 39(9) | 38(8) |
| 4 | 49(14) | 29(5.5) | 49(14) |
| 5 | 58(19) | 46(11) | 56(17.5) |
| 6 | 37(7) | 48(12) | 60(21) |
| 7 | 29(5.5) | 49(14) | 56(17.5) |
| 8 | 44(10) | 65(23) | 62(22) |
| Mean Rank  SD | 9.56  6.25 | 9.94  6.84 | 18.00  5.09 |
| Sum of ranks | 76.5 | 79.5 | 144 |

Find "Tc", the total of the ranks for each group. Just add together all of the ranks for each group in turn.

Here, Tc1 (the rank total for the "no exercise" group) is 76.5.

Tc2 (the rank total for the "20 minutes" group) is 79.5.

Tc3 (the rank total for the "60 minutes" group) is 144.

The degrees of freedom are the number of groups minus one. Here we have three groups, and so we have 2 d.f.

Assessing the significance of *H* depends on the number of participants and the number of groups. If we have more than five participants per group, then treat *H* as Chi-Square. H is statistically significant if it is equal to or larger than the critical value of Chi- Square for your particular d.f. (The table of Chi-Square values).

Here, we have eight participants per group, and so we treat *H* as Chi-Square. *H* is 7.27, with 2 d.f. Here's the relevant part of the Chi-Square table:

Table 3.4 Chi-Square critical value values

|  |  |
| --- | --- |
| *df* | *P=0.05* |
| 1 | 3.84 |
| 2 | 5.99 |
| 3 | 7.82 |

Look along the row that corresponds to our number of degrees of freedom. So in this case, we look along the row for 2 d.f. We compare our obtained value of *H* to each of the critical values in that row of the table, starting on the left hand side and stopping once our value of *H* is no longer equal to or larger than the critical value.

So here, we start by comparing our *H* of 7.27 to 5.99. With 2 degrees of freedom, a value of Chi-Square as large as 5.99 is likely to occur by chance only 5 times in a hundred: i.e. it has a *p* of .05. Our obtained value of 7.27 is even larger than this, and so this tells us that our value of *H* is even *less* likely to occur by chance. Our *H* will occur by chance with a probability of *less* than 0.05.

Kruskal-Wallis test merely tells that the groups differ. We need to inspect the group means or medians to decide precisely how they differ. In proposed case, the interpretation seems fairly straightforward: exercise does seem to reduce self-reported ratings of depression, but only in the case of participants who are doing 40 minutes of it. There seems to be no difference between those participants who took 20 minutes of exercise per day, and those who did not exercise at all.

