Kruskal-Wallis test: rationale

The Kruskal-Wallis test is a nonparametric test in the sense that we do NOT make the assumption that the data were generated from a parametric distribution.

test uses the idea of the so-called 'one-way ANOVA', where inter-class variation is compared to the total variation. In this nonparametric alternative, we consider the variations of the ranks. It is a generalization of the Mann-Whitney-Wilcoxon test.

Basic notation and test statistic

Assume that we have k samples $s_1, ..., s_k$ and let us consider a Kruskal-Wallis test, which has null hypothesis H_0: 'All $k \geq 2$ samples come from the same generating distribution' versus H_1: 'At least two samples have a different generating distribution', or equivalently H_0: 'All samples $k \geq 2$ come from the same population'. Let n_j be the size of the sample s_j, for $j = 1, ..., k$ Let $N = \sum_{j=1}^{k} n_j$. Let S be the pooled sample, with all observations from $s_1, ..., s_k$ and $R = rank(S)$. Finally, let $\bar{R}_j$ be the average rank of the observations from the sample s_j in the pooled sample S. Then, the Kruskal-Wallis test statistic, KW, is given by

$$
\begin{aligned}
KW &= \frac{12}{N(N-1)} \sum_{j=1}^{k} n_j \left(\bar{R}_j - \frac{N+1}{2} \right)^2 \\
&= \frac{12}{N(N-1)} \sum_{j=1}^{k} n_j \bar{R}_j^2 - 3(N+1)
\end{aligned}
$$

Critical region

The distribution of the test statistic KW under H_0 depends on the n_j's and is thus difficult to find in tables. However, the distribution of KW can be simulated from (see example).

The critical region R_α has the form

$$R_\alpha = \left\{ q_{1-\alpha}, ..., c \right\}$$

where $q_{1-\alpha}$ is the $(1-\alpha)$-quantile of the distribution of KW under H_0 and c is the maximum value that can take the test statistic.

Asymptotic distribution of the test statistic

When the n_j's are large, then, under H_0, the test statistic KW has the following distribution asymptotically

$$KW \sim \chi^2_{(k-1)}$$

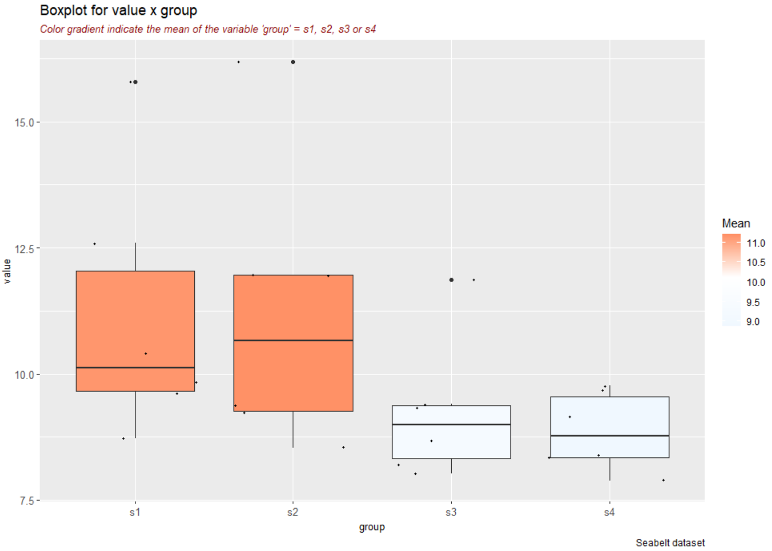
H_0 is then rejected with an approximate α level of significance if $KW > \chi^2_{1-\alpha,(k-1)}$.

Example 1

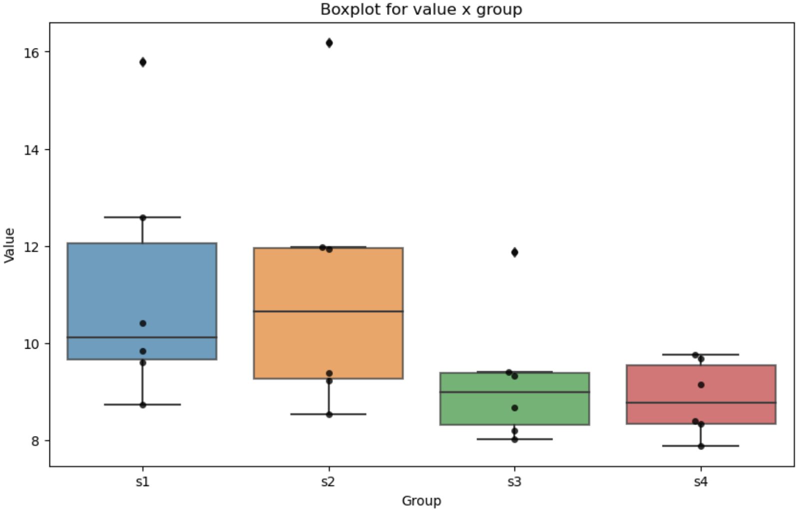
Example: With the data of the Seabelts dataset of R, we compute casualty rates for the first semesters of years 1978 to 1981. The data are shown in the table below.Let's first perform the Kruskal-Wallis test. Then, we can simulate from the exact distribution of the test statistic under H_0 and determine if H_0 is rejected for those data at $\alpha = 5\%$ level of significance.

i	s_1	s_2	s_3	s_4
1	15.791	16.193	11.870	9.681
2	12.595	11.937	9.400	9.764
3	10.405	11.968	9.322	9.154
4	9.836	9.376	8.200	8.330
5	8.729	9.227	8.020	8.388
6	9.608	8.539	8.671	7.888

Visualizing the data in R



Visualizing the data in Python



Working example

The ranks R_j in the pooled sample and the average ranks $\bar{R}_j$ are given in the following tables.

i	R_1	R_2	R_3	R_4
1	23	24	19	15
2	22	20	13	16
3	18	21	11	9
4	17	12	3	4
5	8	10	2	5
6	14	6	7	1

$\bar{R}_1$	$\bar{R}_2$	$\bar{R}_3$	$\bar{R}_4$
17	15.5	9.17	8.33

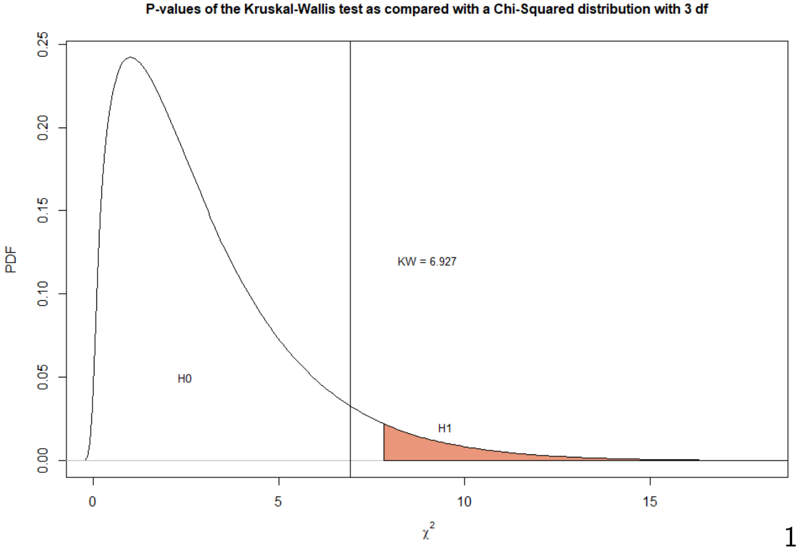
Test statistic and test decision

In addition, we have $n_1 = n_2 = n_3 = n_4 = 6$ and N $= \sum_{j=1}^{k} n_j = 24$. The value of our test statistic KW is thus

$$\frac{12}{N(N-1)} \sum_{j=1}^{k} n_j \left(\bar{R}_j - \frac{N+1}{2}\right)^2$$
$$= \left(\frac{12}{24(24-1)}\right) \left[6(17-12.5)^2 + 6(15.5-12.5)^2 + 6(9.17-12.5)^2 + 6(8.33-12.5)^2\right]$$
$$= 6.927$$

Since the 0.95-quantile of a $\chi_{(3)} = 7.815$, we can NOT reject H_0 in this case.

Visualizing the test decision



Kruskal-Wallis test in R

```r
k=4  # number of samples

# data
s1 <- c(15.791,12.595,10.405,9.836,8.729,9.608)
s2 <- c(16.193,11.937,11.968,9.376,9.227,8.539)
s3 <- c(11.870,9.400, 9.322, 8.200,8.020,8.671)
s4 <- c(9.681, 9.764, 9.154, 8.330,8.388,7.888)

# total sample size N = sum nj
N <- length(s1)+length(s2)+length(s3)+length(s4)

# matrix containing the ranks Rj in the pooled sample
Rk <- matrix(rank(c(s1,s2,s3,s4)), ncol = k, byrow = FALSE) # vector of ranks in pooled sample

# mean of the Rj
Rbar <- apply(Rk,2,mean)
Rbar  # [1] 17.00 15.50  9.17  8.33 (R_bar_j)

# test statistic
KW <- 12/(N*(N+1)) * sum( 6 * ( Rbar - (N+1)/2   )^2 )
KW  # [1] 6.926667

# Tf the sample sizes are large, under H0, K has approximately a chi-squared(k-1) distribution,
# where k is the number of samples.
qchisq(p = 0.95 , df = 3)  # [1] 7.814728
# the 95% quantile is 7.814728 and so K = 6.926667 is not in
# the critical region if we take alpha = 5%.
# Thus we don't reject H0.
```

Kruskal-Wallis test in Python

```
import numpy as np
import pandas as pd
from scipy.stats import rankdata
from scipy.stats import chi2

k = 4 # number of samples
# data
s1 = [15.791, 12.595, 10.405, 9.836, 8.729, 9.608]
s2 = [16.193, 11.937, 11.968, 9.376, 9.227, 8.539]
s3 = [11.870, 9.400, 9.322, 8.200, 8.020, 8.671]
s4 = [9.681, 9.764, 9.154, 8.330, 8.388, 7.888]

# Combine all samples into a single array
all_samples = np.concatenate([s1, s2, s3, s4])
all_samples = np.transpose(all_samples.reshape(4,6))

# Assign ranks to the data in the pooled sample
ranks = rankdata(all_samples)
ranks.reshape(6,4)
Rbar = np.mean(Rk, axis=0)
N = 4*len(all_samples)

# Calculate the test statistic for the Kruskal-Wallis test
KW = 12/(N*(N+1)) * np.sum(6 * (Rbar - (N+1)/2)**2)

df = 3
alpha = 0.05
critical_value = chi2.ppf(1 - alpha, df)
# the 95% quantile is 7.814728 and so KW = 6.926667 is not in
# the critical region if we take alpha = 5%.
# Thus we don't reject H0.
```

Kruskal-Wallis test with inbuilt functions in R and Python

In the R language:

```r
# Kruskal-Wallis test with inbuilt function
kruskal.test(list(s1,s2,s3,s4))

# Kruskal-Wallis rank sum test
# data:  list(s1, s2, s3, s4)
# Kruskal-Wallis chi-squared = 6.9267, df = 3, p-value = 0.07427
```

And in Python:

```python
from scipy.stats import kruskal

# Combine the data into a list of arrays
data_list = [s1, s2, s3, s4]
# Perform the Kruskal-Wallis test
result = kruskal(*data_list)

# Print the result
print("Kruskal-Wallis test statistic:", result.statistic)
print("p-value:", result.pvalue)
Kruskal-Wallis test statistic: 6.926666666666662
p-value: 0.07427225132882118
```

References

Bagdonavičius V., Kruopis J., Nikulin M. S., Non-parametric Tests for Complete Data (2011), Wiley, ISBN 978-1-84821-269-5 (hardback)

The R Project for Statistical Computing:
https://www.r-project.org/

Python:
https://www.python.org/

course notes