Title

stata.com

kwallis — Kruskal–Wallis equality-of-populations rank test

Syntax

`kwallis` *varname* [*if*] [*in*] , `by(`*groupvar*`)`

Menu

Statistics > Nonparametric analysis > Tests of hypotheses > Kruskal-Wallis rank test

Description

`kwallis` tests the hypothesis that several samples are from the same population. In the syntax diagram above, *varname* refers to the variable recording the outcome, and *groupvar* refers to the variable denoting the population. `by()` is required.

Option

`by(`*groupvar*`)` is required. It specifies a variable that identifies the groups.

Remarks and examples

stata.com

▷ Example 1

We have data on the 50 states. The data contain the median age of the population, `medage`, and the region of the country, `region`, for each state. We wish to test for the equality of the median age distribution across all four regions simultaneously:

```
. use http://www.stata-press.com/data/r13/census
(1980 Census data by state)
. kwallis medage, by(region)
Kruskal-Wallis equality-of-populations rank test
```

region	Obs	Rank Sum
NE	9	376.50
N Cntrl	12	294.00
South	16	398.00
West	13	206.50

```
chi-squared =     17.041 with 3 d.f.
probability =      0.0007

chi-squared with ties =     17.062 with 3 d.f.
probability =      0.0007
```

From the output, we see that we can reject the hypothesis that the populations are the same at any level below 0.07%.

◁

Stored results

`kwallis` stores the following in `r()`:

Scalars

`r(df)`	degrees of freedom
`r(chi2)`	χ^2
`r(chi2_adj)`	χ^2 adjusted for ties

Methods and formulas

The Kruskal–Wallis test (Kruskal and Wallis 1952, 1953; also see Altman [1991, 213–215]; Conover [1999, 288–297]; and Riffenburgh [2012, sec. 11.6]) is a multiple-sample generalization of the two-sample Wilcoxon (also called Mann–Whitney) rank sum test (Wilcoxon 1945; Mann and Whitney 1947). Samples of sizes n_j, $j = 1, \dots, m$, are combined and ranked in ascending order of magnitude. Tied values are assigned the average ranks. Let n denote the overall sample size, and let $R_j = \sum_{i=1}^{n_j} R(X_{ji})$ denote the sum of the ranks for the jth sample. The Kruskal–Wallis one-way analysis-of-variance test, H, is defined as

$$H = \frac{1}{S^2}\left\{\sum_{j=1}^{m}\frac{R_j^2}{n_j} - \frac{n(n+1)^2}{4}\right\}$$

where

$$S^2 = \frac{1}{n-1}\left\{\sum_{\text{all ranks}} R(X_{ji})^2 - \frac{n(n+1)^2}{4}\right\}$$

If there are no ties, this equation simplifies to

$$H = \frac{12}{n(n+1)}\sum_{j=1}^{m}\frac{R_j^2}{n_j} - 3(n+1)$$

The sampling distribution of H is approximately χ^2 with $m-1$ degrees of freedom.

William Henry Kruskal (1919–2005) was born in New York City. He studied mathematics and statistics at Antioch College, Harvard, and Columbia, and joined the University of Chicago in 1951. He made many outstanding contributions to linear models, nonparametric statistics, government statistics, and the history and methodology of statistics.

Wilson Allen Wallis (1912–1998) was born in Philadelphia. He studied psychology and economics at the Universities of Minnesota and Chicago and at Columbia. He taught at Yale, Stanford, and Chicago, before moving as president (later chancellor) to the University of Rochester in 1962. He also served in several Republican administrations. Wallis served as editor of the *Journal of the American Statistical Association*, coauthored a popular introduction to statistics, and contributed to nonparametric statistics.

References

Altman, D. G. 1991. *Practical Statistics for Medical Research*. London: Chapman & Hall/CRC.

Conover, W. J. 1999. *Practical Nonparametric Statistics*. 3rd ed. New York: Wiley.

Fienberg, S. E., S. M. Stigler, and J. M. Tanur. 2007. The William Kruskal Legacy: 1919–2005. *Statistical Science* 22: 255–261.

Kruskal, W. H., and W. A. Wallis. 1952. Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association* 47: 583–621.

——. 1953. Errata: Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association* 48: 907–911.

Mann, H. B., and D. R. Whitney. 1947. On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics* 18: 50–60.

Newson, R. B. 2006. Confidence intervals for rank statistics: Somers' D and extensions. *Stata Journal* 6: 309–334.

Olkin, I. 1991. A conversation with W. Allen Wallis. *Statistical Science* 6: 121–140.

Riffenburgh, R. H. 2012. *Statistics in Medicine*. 3rd ed. San Diego, CA: Academic Press.

Wilcoxon, F. 1945. Individual comparisons by ranking methods. *Biometrics* 1: 80–83.

Zabell, S. L. 1994. A conversation with William Kruskal. *Statistical Science* 9: 285–303.

Also see

[R] **nptrend** — Test for trend across ordered groups

[R] **oneway** — One-way analysis of variance

[R] **sdtest** — Variance-comparison tests

[R] **signrank** — Equality tests on matched data