snp16.*y*	Robust confidence intervals for median (and other percentile) differences between two groups

Author: Roger Newson, King's College London, UK. Email: roger.newson@kcl.ac.uk Date: 12 August 2005.

Abstract: A program is presented for calculating robust confidence intervals for Hodges-Lehmann median (and other percentile) differences (and ratios) between values of a variable in two samples. The median difference is the same as that produced by the programs `cid` and `npshift`, using the Conover method. However, the confidence limits are typically different, being robust to the possibility that the two population distributions differ in ways other than location, such as having unequal variances. The program uses the program `somersd`, and is part of the `somersd` package.

Keywords: Robust, confidence interval, median, percentile, difference, ratio, rank-sum, Wilcoxon, two-sample.

Syntax

`cendif` *depvar* [`using` *filename*][*weight*][`if` *exp*][`in` *range*], `by(`*groupvar*`)` [`centile(`*numlist*`)` `level(`#`)` `eform` `cluster(`*varname*`)` `tdist` `transf(`*transformation_name*`)` `saving(`*filename*[`,replace`]`)` `nohold`]

where *transformation_name* is one of

`iden | z | asin`

`fweights`, `iweights` and `pweights` are allowed; see [U] **14.1.6 weight**. They are treated as described in **Methods and Formulas** below.

Description

`cendif` calculates confidence intervals for Hodges-Lehmann median differences, and other percentile differences, between values of a Y-variable in *depvar* for a pair of observations chosen at random from two groups A and B, defined by the *groupvar* in the `by` option. These confidence intervals are robust to the possibility that the population distributions in the two groups are different in ways other than location. This might happen if, for example, the two populations had different variances. For positive-valued variables, `cendif` can be used to calculate confidence intervals for median ratios or other percentile ratios. `cendif` is part of the `somersd` package, and requires the program `somersd` (Newson, 2000a) in order to work.

Options

`by(`*groupvar*`)` is not optional. It specifies the name of the grouping variable. This variable must have exactly two possible values. The lower value indicates Group A, and the higher value indicates Group B.

`centile(`*numlist*`)` specifies a list of percentile differences to be reported, and defaults to `centile(50)` (median only) if not specified. Specifying `centile(25 50 75)` will produce the 25th, 50th and 75th percentile differences.

`level(`# `)` specifies the confidence level (percent) for confidence intervals; see help for `level`.

`eform` specifies that exponentiated percentile differences are to be given. This option is used if *depvar* is the log of a positive-valued variable. In this case, confidence intervals are calculated for percentile ratios between values of the original positive variable, instead of for percentile differences.

`cluster(`*varname*`)` specifies the variable which defines sampling clusters. If `cluster` is defined, then the percentiles are calculated using the between-cluster Somers' D, and the confidence intervals are calculated assuming that the data are a sample of clusters from a population of clusters, rather than a sample of observations from a population of observations.

`tdist` specifies that the standardized Somers' D estimates are assumed to be sampled from a t-distribution with n-1 degrees of freedom, where n is the number of clusters, or the number of observations if `cluster` is not specified.

`transf(`*transformation_name*`)` specifies that the Somers' D estimates are to be transformed, defining a standard error for the transformed population value, from which the confidence limits for the percentile differences are calculated. `z` (the default) specifies Fisher's z (the hyperbolic arctangent), `asin` specifies Daniels' arcsine, and `iden` specifies identity or untransformed.

`saving(`*filename*`[,replace])` specifies a data set, to be created, whose observations correspond to the observed values of differences between a value of *depvar* in Group A and a value of *depvar* in Group B. `replace` instructs Stata to replace any existing data set of the same name. The saved data set can then be re-used if `cendif` is called later, with `using`, to save the large amounts of processing time used to calculate the set of observed differences. The `saving` option and the `using` utility are provided mainly for programmers to use, at their own risk.

`nohold` indicates that any existing estimation results are to be overwritten with a new set of estimation results, for the use of programmers. In default, any existing estimation results are restored after execution of `cendif`.

Methods and Formulas

Suppose that a population contains two disjoint sub-populations A and B, and a random variable Y is defined for individuals from both sub-populations. For $0 < q < 1$, a $100q$th percentile difference in Y between Populations A and B is defined as a value θ satisfying

$$D[Y^*(\theta)|X] = 1 - 2q, \tag{1}$$

where X is a binary variable equal to 1 for Population A and 0 for Population B, $Y^*(\theta)$ is defined as Y if $X = 1$ and $Y + \theta$ if $X = 0$, and $D[\,\cdot\,|\,\cdot\,]$ denotes Somers' D (Somers, 1962; Newson, 2000). Somers' D is defined as

$$D[V|W] = E\left[\,\mathrm{sign}(V_1 - V_2)\,\mathrm{sign}(W_1 - W_2)\,\right] / E\left[\mathrm{sign}(W_1 - W_2)^2\right], \tag{2}$$

where (W_1, V_1) and (W_2, V_2) are bivariate data points sampled independently from the same population, and $E[\,\cdot\,]$ denotes expectation. In the case of (1), where $W = X$ and $V = Y^*(\theta)$, Somers' D is the difference between two conditional probabilities. Given an individual sampled from Population A and an individual sampled from Population B, these are the probability that the individual from Population A has the higher Y^*-value and the probability that the individual from Population B has the higher Y^*-value. Somers' D is therefore the parameter equal to zero under the null hypothesis tested by the "non-parametric" Wilcoxon rank-sum test on $Y^*(\theta)$. In the case where $q = 0.5$ (and therefore $1 - 2q = 0$), a $100q$th percentile difference is known as a Hodges-Lehmann median percentile difference, and is zero under the null hypothesis tested by a Wilcoxon rank-sum test on Y. The median percentile difference was introduced explicitly by Hodges and Lehmann (1963), but is also a special case of the Theil median slope (Theil, 1950) for a binary X-variable.

Note that a value of θ satisfying (1) is not always unique. If Y has a discrete distribution, then there may be no solution, or a wide interval of solutions. However, the method used here is intended to produce a confidence interval containing any given θ satisfying (1), with a probability at least equal to the confidence level, if such a θ exists.

We will assume that there are N_1 observations sampled from Population A and N_2 observations sampled from Population B, giving a total of $N_1 + N_2 = N$ observations. These observations will be identified by double subscripts, so that Y_{ij} is the Y-value for the jth observation sampled from the ith population (where i=1 for Population A and $i = 2$ for Population B). The corresponding X-values (ones and zeros) will be denoted X_{ij}. The observations will be assumed to have importance weights (`iweights` or `pweights`) denoted w_{ij}, and cluster sequence numbers denoted c_{ij}. `cendif` follows the usual Stata practice of assuming an `fweight` to stand for multiple observations, with the same values for all other variables. The clusters may be nested within the two groups or contain observations from each of the two groups, but the percentile differences will only apply to observations from distinct clusters. If clusters are present, then the confidence intervals will be calculated assuming that the sample was generated by sampling clusters independently from a population of clusters, rather than by sampling N observations independently from the total population of observations, or by sampling N_1 and N_2 observations from Populations A and B, respectively. (In default, all the w_{ij} will be ones, and the c_{ij} will be in sequence from 1 to N, so the difference between these three alternatives will not matter.) We will denote by M the number of distinct values of a difference $Y_{1j} - Y_{2k}$ observed between Y-values in the two samples belonging to different clusters. The difference values themselves will be denoted $t_1, \ldots, t_M$. For each h from 1 to M, we define the sum of product weights of differences equal to t_h as

$$W_h = \sum_{j,k:\, Y_{1j} - Y_{2k} = t_h} \delta(c_j, c_k) w_{1j} w_{2k}, \tag{3}$$

where $\delta(a, b)$ is 0 if $a = b$ and 1 if $a \neq b$. Given a value of θ expressed in units of Y, we can define $Y^*_{ij}(\theta)$ to be Y_{ij} if $i = 1$, and $Y_{ij} + \theta$ if $i = 2$. The sample Somers' D of $Y^*(\theta)$ with respect to X is defined as

$$D^*(\theta) = \hat{D}[Y^*(\theta)|X] = \frac{\sum_{j=1}^{N_1} \sum_{k=1}^{N_2} \delta(c_{1j}, c_{2k})\, w_{1j} w_{2k}\, \mathrm{sign}(Y_{1j} - Y_{2k} - \theta)}{\sum_{j=1}^{N_1} \sum_{k=1}^{N_2} \delta(c_{1j}, c_{2k})\, w_{1j} w_{2k}} = \frac{\sum_{h:\, t_h > \theta} W_h - \sum_{h:\, t_h < \theta} W_h}{\sum_{h=1}^{M} W_h}, \tag{4}$$

where $\hat{D}[\,\cdot\,|\,\cdot\,]$ denotes the sample Somers' D, defined by the methods of Newson (2000). Clearly, given a sample, $D^*(\theta)$ is a nonincreasing function of θ. (Note that only between-cluster differences are included.) Figure 1 shows $D^*(\theta)$ as a function of θ for differences between trunk capacities of US and foreign cars (expressed in cubic feet) in the `auto` data. The squares represent the values $D^*(t_h)$ for the observed differences t_h. Note that $D^*(\theta)$ is discontinuous at the observed differences, and constant in each open interval between two successive observed differences.

We aim to include θ in a confidence interval for a qth percentile difference if, and only if, the sample $D^*(\theta)$ is compatible with a *population* $D[Y^*(\theta)|X]$ equal to $1-2q$. The methods of Newson (2000), used by the program `somersd`, typically use a transformation $\zeta(\cdot)$, which, for present purposes, may either be the identity, the arscine or Fishers' z (the hyperbolic arctangent). The transformed sample statistic $\hat{\zeta}(\theta) = \zeta[D^*(\theta)]$ is assumed to be Normally distributed around the population parameter $\zeta\{D[Y^*(\theta)|X]\}$. In the present application, we assume that, if $D[Y^*(\theta)|X] = 1-2q$, then the quantity

$$[\,\hat{\zeta}(\theta) - \zeta(1-2q)\,] \,/\, \mathrm{SE}[\hat{\zeta}(\theta)] \tag{5}$$

has a standard Normal distribution, where $\mathrm{SE}[\hat{\zeta}(\theta)]$ is the sampling standard deviation (or standard error) of $\zeta[D^*(\theta)]$. If we knew the value of $\mathrm{SE}[\hat{\zeta}(\theta)]$, then a $100(1-\alpha)\%$ confidence interval for a qth percentile difference might be the interval of values of θ for which

$$\zeta^{-1}\{\,\zeta(1-2q) - z_\alpha\,\mathrm{SE}[\hat{\zeta}(\theta)]\,\} \leq D^*(\theta) \leq \zeta^{-1}\{\,\zeta(1-2q) + z_\alpha\,\mathrm{SE}[\hat{\zeta}(\theta)]\,\}, \tag{6}$$

where z_α is the $100(1-\frac{1}{2}\alpha)$th percentile of the standard Normal distribution.

To construct such a confidence interval, we proceed as follows. Given a value of D, define

$$B_\mathrm{L}(D) = \inf\{\theta : D^*(\theta) \leq D\}, \quad B_\mathrm{R}(D) = \sup\{\theta : D^*(\theta) \geq D\},$$

$$B_\mathrm{C}(D) = \begin{cases} B_\mathrm{L}(D), & \text{if } B_\mathrm{R}(D) = \infty \\ B_\mathrm{R}(D), & \text{if } B_\mathrm{L}(D) = -\infty \\ [\,B_\mathrm{L}(D) + B_\mathrm{R}(D)\,]/2, & \text{otherwise.} \end{cases} \tag{7}$$

(By convention, the supremum (or infimum) of a set unbounded to the right (or left) are defined as ∞ and $-\infty$, respectively.) Clearly, $B_\mathrm{L}(D) \leq B_\mathrm{C}(D) \leq B_\mathrm{R}(D)$, and the values of $B_\mathrm{L}(D)$ and $B_\mathrm{R}(D)$ (if finite) can be either the same t_h, or two successive ones. The confidence interval for the qth percentile difference is centred on the sample qth percentile difference

$$\hat{\xi}_q = B_\mathrm{C}(1-2q). \tag{8}$$

`cendif` then calls `somersd`, with the X_{ij} as the predictor variable, and the $Y^*_{ij}(\hat{\xi}_q)$, for the values of q implied by the `centile` option, as the predicted variables. The standard errors generated by `somersd` are used as estimates $\widehat{\mathrm{SE}}[\hat{\zeta}(\hat{\xi}_q)]$ of the standard error of $\hat{\zeta}(\theta)$ where θ satisfies (1). The lower and upper confidence limits for the qth percentile difference are, respectively,

$$\hat{\xi}_q^{(\min)} = B_\mathrm{L}\Big(\zeta^{-1}\{\,\zeta(1-2q) - z_\alpha\,\widehat{\mathrm{SE}}[\hat{\zeta}(\hat{\xi}_q)]\,\}\Big), \quad \hat{\xi}_q^{(\max)} = B_\mathrm{R}\Big(\zeta^{-1}\{\,\zeta(1-2q) + z_\alpha\,\widehat{\mathrm{SE}}[\hat{\zeta}(\hat{\xi}_q)]\,\}\Big). \tag{9}$$

If `tdist` is specified, then `cendif` uses the t-distribution with $N-1$ degrees of freedom (or $N_\mathrm{clust}-1$ degrees of freedom if there are N_clust clusters) instead of the normal distribution, so t_α replaces z_α in (6) and (9). Note that the upper and lower confidence limits may occasionally be infinite, in the case of extreme percentiles and/or very small sample numbers. (`cendif` codes these infinite limits as plus or minus the "magic number" `1E+300`, or $\pm 10^{300}$.) Figure 1 shows the median difference in trunk capacity, and its confidence limits, as reference lines on the horizontal axis. The estimated median difference is 3 cubic feet, with 95% confidence limits from 1 to 5 cubic feet. The reference lines on the vertical axis are the optimum, minimum and maximum values of $D^*(\theta)$ required for θ to be in the confidence interval. These values of $D^*(\theta)$ are saved by `cendif` in the matrix `r(Dsmat)`. If the option `saving` is specified, then `cendif` also saves an output data set with M observations corresponding to the ordered differences t_h. The variables are `diff` (containing the t_h), `weight` (containing the W_h), `Dstar` (containing the $D^*(t_h)$), and `Dstar_r`, which contains the right-hand limiting value of $D^*(\theta)$,

$$D^*_R(t_h) = \lim_{\theta \to t_h +} D^*(\theta), \tag{10}$$

which is the value of $D^*(\theta)$ in the open interval (t_h, t_{h+1}) for $h < M$.

Conover (1980) presents a method which, for large samples, is essentially equivalent to (6), in the special case where $q = 0.5$ and $\zeta(D) = D$. (This is the method for calculating confidence intervals for median differences popularized by Campbell and Gardner (1988) and Gardner and Altman (1989), and available in Stata using Duolao Wang's `npshift` routine (Wang, 1999) or Patrick Royston's `cid` routine, downloadable from SSC (Royston, 1998).) However, Conover's method uses the assumption that the two population distributions are different only in location.

This assumption (essentially) enables the calculation of SE$[\hat{\zeta}(\theta)]$ for large samples, and of the exact distribution of $D^*(\theta)$ for small samples. It also implies that the median difference is the difference between medians. In the present case, we are not making this assumption, as the confidence interval is intended to be robust to the possibility that the two populations are different in ways other than location. (For instance, the two populations might be unequally variable.) The median difference is therefore not necessarily the difference between medians. Also, we have to estimate SE$[\hat{\zeta}(\theta)]$, and this estimate is itself subject to some amount of sampling error. The method of `cendif` compares to Conover's method as the unequal-variance t-test compares to the equal-variance t-test. Conover's method, like the equal-variance t-test, assumes that you can use data from the larger of two samples to estimate the population variability of the smaller sample.

I have been carrying out some simulations of sampling from two Normal populations, with a view to finding the coverage probabilities and geometric mean lengths of the confidence intervals for the median difference generated by `cid` and by `cendif` with the `tdist` option. So far, I find that, even with small sample sizes, the `cendif` method consistently gives coverage probabilities closer to the nominal value than the Conover method when variances are unequal, in which case `cid` produces confidence intervals either too wide or too narrow, depending on whether the larger or smaller sample has the greater population variance. Usually, the difference in coverage probability is small (1% or 2%), so the Conover method performs fairly well, in spite of false assumptions. However, if a sample of 20 is compared to a sample of 10, and the population standard deviation of the smaller sample is three times that of the larger sample, then the nominal 95% confidence interval has a true coverage probability of only 90% under the Conover method, compared to 94% under the `cendif` method. The two methods show little or no difference, either in geometric mean confidence interval width or in coverage probability, when the variances are equal and the Conover assumption is therefore true. From the results so far, I would therefore recommend the `cendif` method as an improved version of the Conover method, offering insurance against the possibility that the Conover assumption is wildly wrong, at little or no price in performance if the Conover assumption is right. However, I hope to carry out further simulations on the two methods, and to report the results in due course.

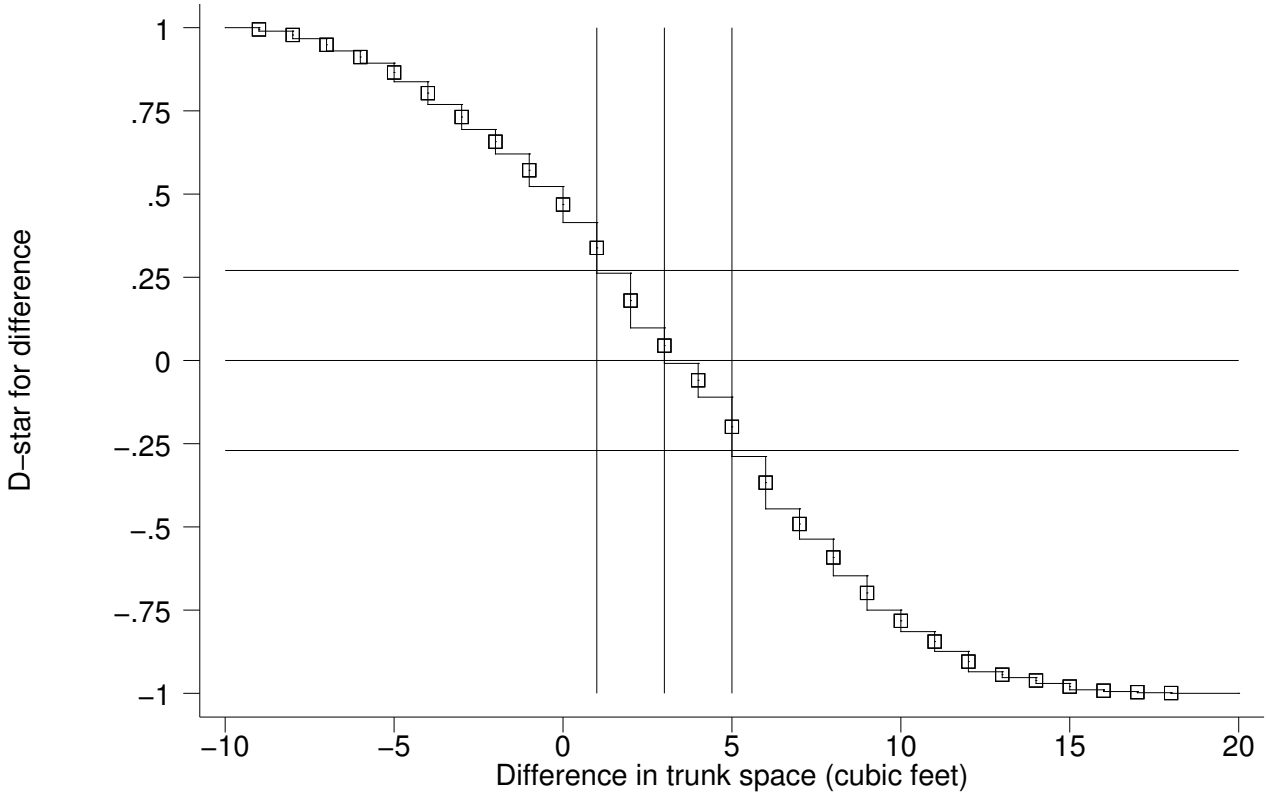

Figure 1. $D^*(\theta)$ plotted against the difference θ in trunk space between US and foreign cars

Example 1

In the `auto` data, we compare weights of US cars and foreign cars. We use `cid` and `cendif` to estimate the median difference:

```
. cid weight,by(foreign) median unpaired
Rank-based confidence interval for difference in  medians by foreign
Variable |     Obs     Estimate           K        [95% Conf. Interval]
---------+-------------------------------------------------------------
  weight |      74         1095         406             720        1350
. cendif weight,by(foreign)
Y-variable: weight (Weight (lbs.))
Grouped by: foreign (Car type)
Group numbers:
   Car type |      Freq.     Percent        Cum.
------------+-----------------------------------
   Domestic |         52       70.27       70.27
    Foreign |         22       29.73      100.00
------------+-----------------------------------
      Total |         74      100.00
Transformation: Fisher's z
95% confidence interval(s) for percentile difference(s)
between values of weight in first and second groups:
     Percent  Pctl_Dif   Minimum   Maximum
r1        50      1095       750      1330
```

We note that the median difference in weight is 1095 pounds, according to both `cid` and `cendif`. However, the confidence limits given by `cendif` are 750 and 1330 pounds, whereas the confidence limits given by `cid` are 720 and 1350 pounds. This is because foreign cars are fewer in number, and less variable in weight, than US cars, and `cid` assumes equal variances, whereas `cendif` allows for unequal variances. If we carry out equal-variance and unequal-variance t-tests (not shown), we find a similar difference in the width of the confidence limits for the mean difference.

`cendif` can also calculate confidence intervals for percentiles other than medians. These contain information about the degree of overlap between the two populations. Here, we estimate the 25th, 50th and 75th percentile differences, using the `centile` option:

```
. cendif weight,by(foreign) ce(25 50 75)
Y-variable: weight (Weight (lbs.))
Grouped by: foreign (Car type)
Group numbers:
   Car type |      Freq.     Percent        Cum.
------------+-----------------------------------
   Domestic |         52       70.27       70.27
    Foreign |         22       29.73      100.00
------------+-----------------------------------
      Total |         74      100.00
Transformation: Fisher's z
95% confidence interval(s) for percentile difference(s)
between values of weight in first and second groups:
     Percent  Pctl_Dif   Minimum   Maximum
r1        25       485       100       810
r2        50      1095       750      1330
r3        75      1555      1320      1790
```

If we want to estimate percentile ratios of weight, rather than percentile differences, then we simply take logs and use the `eform` option:

```
. gene logwt=log(weight)
. cendif logwt,by(foreign) ce(25 50 75) eform
Y-variable: logwt
Grouped by: foreign (Car type)
Group numbers:
   Car type |      Freq.     Percent        Cum.
------------+-----------------------------------
   Domestic |         52       70.27       70.27
    Foreign |         22       29.73      100.00
------------+-----------------------------------
      Total |         74      100.00
Transformation: Fisher's z
95% confidence interval(s) for percentile ratio(s)
between values of exp(logwt) in first and second groups:
      Percent   Pctl_Rat    Minimum    Maximum
r1         25  1.1935375  1.0341465  1.3533567
r2         50  1.4806389  1.3101849  1.6280196
r3         75   1.744916  1.6079542  1.8772724
```

We note that, typically, US cars are 148% as heavy as foreign cars, with confidence limits ranging from 131% to 163% as heavy. The 25th percentile ratio (103% to 135%) shows that the two car types do not overlap a great deal.

Acknowledgements

I would like to thank Nicholas J. Cox of Durham University, UK, and William Gould of StataCorp for some very helpful advice on the coding of infinite confidence limits, such as those occasionally resulting from Equation (9).

Saved results

`cendif` saves in `r()`:

Scalars

`r(N)`	number of observations	`r(N_clust)`	number of clusters
`r(N_1)`	sample size N_1	`r(N_2)`	sample size N_2
`r(df_r)`	residual degrees of freedom (if `tdist` present)		

Macros

`r(depvar)`	name of *Y*-variable	`r(by)`	name of `by` variable defining groups
`r(clustvar)`	name of cluster variable	`r(tdist)`	`tdist` if specified
`r(wtype)`	weight type	`r(wexp)`	weight expression
`r(centiles)`	list of percents for percentiles	`r(Dslist)`	list of D^*-values for percentiles
`r(transf)`	transformation specified by `transf`	`r(tranlab)`	transformation label in output
`r(level)`	confidence level	`r(eform)`	`eform` if specified

Matrices

`r(cimat)`	confidence intervals for differences or ratios	`r(Dsmat)`	upper and lower limits for $D^*(\theta)$

Historical note

This document is a post-publication update of an article which appeared in the Stata Technical Bulletin (STB) as Newson (2000d). The `somersd` package appeared in Newson (2000a), and a post-publication update of that STB article is distributed with this document as part of the documentation of the `somersd` package. The `somersd` package was later revised in Newson (2000b), Newson (2000c), Newson (2000d), Newson (2001a) and Newson (2001b). After 2001, STB was replaced by The Stata Journal (SJ), and all subsequent updates to the `somersd` package only appeared on SSC and on Roger Newson's homepage at *http://www.kcl-phs.org.uk/rogernewson*, which is accessible from within net-aware Stata. However, Newson (2002) gives a comprehensive review of Somers' D, Kendall's τ_a, median differences, and their estimation in Stata using the `somersd` package.

References

Campbell, M. J. and M. J. Gardner. Calculating confidence intervals for some non-parametric analyses. *British Medical Journal* 296: 1454-1456.

Conover, W. J. 1980. *Practical Nonparametric Statistics*. 2d ed. New York: John Wiley & Sons.

Gardner, M. J. and D. G. Altman. 1989. *Statistics with Confidence*. London: British Medical Journal.

Hodges, J. L. and E. L. Lehmann. 1963. Estimates of location based on rank tests. *Annals of Mathematical Statistics* 34: 598–611.

Newson, R. 2000a. snp15: `somersd` – Confidence intervals for nonparametric statistics and their differences. *Stata Technical Bulletin* 55: 47–55. Reprinted in *Stata Technical Bulletin Reprints*, vol. 10, pp. 312–322.

Newson, R. 2000b. snp15.1: Update to `somersd`. *Stata Technical Bulletin* 57: 35. Reprinted in *Stata Technical Bulletin Reprints*, vol. 10, pp. 322–323.

Newson, R. 2000c. snp15.2: Update to `somersd`. *Stata Technical Bulletin* 58: 30. Reprinted in *Stata Technical Bulletin Reprints*, vol. 10, p. 323.

Newson, R. 2000d. snp16: Robust confidence intervals for median and other percentile differences between groups. *Stata Technical Bulletin* 58: 30–35. Reprinted in *Stata Technical Bulletin Reprints*, vol. 10, pp. 324–331.

Newson, R. 2001a. snp15.3: Update to `somersd`. *Stata Technical Bulletin* 61: 22. Reprinted in *Stata Technical Bulletin Reprints*, vol. 10, p. 33x.

Newson, R. 2001b. snp16.1: Update to `cendif`. *Stata Technical Bulletin* 61: 22. Reprinted in *Stata Technical Bulletin Reprints*, vol. 10, p. 33x.

Newson, R. 2002. Parameters behind "nonparametric" statistics: Kendall's tau, Somers' D and median differences. *The Stata Journal* 2: 45–64. A pre-publication draft can be downloaded from Roger Newson's website at `http://www.kcl-phs.org.uk/rogernewson` using the `net` command in Stata.

Royston, P. 1998. CID: Stata module to calculate confidence intervals for means or differences. On the Ideas list at `http://ideas.uqam.ca/ideas/data/Softwares/bocbocodeS338001.html` as of 30 May 2005.

Somers, R. H. 1962. A New Asymmetric Measure of Association for Ordinal Variables. *American Sociological Review* 27: 799-811.

Theil, H. 1950. A rank invariant method of linear and polynomial regression analysis, I, II, III. *Proceedings of the Koninklijke Nederlandse Akademie Wetenschappen*, *Series A – Mathematical Sciences* 53: 386–392, 521–525, 1397–1412.

Wang, D. 1999. sg123: Hodges-Lehmann estimation of a shift in location between two populations. *Stata Technical Bulletin* 52: 52–53. Reprinted in *Stata Technical Bulletin Reprints*, vol. 9, pp. 255–257.