Robust confidence intervals for Hodges-Lehmann median differences
A simulation study

Roger B. Newson
r.newson@imperial.ac.uk
http://www.imperial.ac.uk/nhli/r.newson/

National Heart and Lung Institute
Imperial College London

13th UK Stata Users’ Group Meeting, 10–11 September, 2007
Downloadable from the conference website at
http://ideas.repec.org/s/boc/usug07.html
What is a Hodges–Lehmann median difference?

- A Theil–Sen median slope of $Y$ with respect to $X$ is a solution in $\beta$ to the equation $D(Y - \beta X | X) = 0$, where $D(\cdot | \cdot)$ denotes the rank association measure Somers’ $D$.

- *In other words*, a median slope is a linear effect of $X$ on $Y$, large enough to explain the observed association.

- If $X$ is binary with values 0 and 1, then the Theil–Sen median slope is the **Hodges–Lehmann median difference** between the subpopulations in which $X = 1$ and $X = 0$.

- *In other words*, the Hodges–Lehmann median difference is the median pairwise difference between two $Y$–values, sampled at random from the two subpopulations.

- Note that the median difference is *not* always the difference between the two subpopulation medians!
What is a Hodges–Lehmann median difference?

- A Theil–Sen median slope of $Y$ with respect to $X$ is a solution in $\beta$ to the equation $D(Y - \beta X | X) = 0$, where $D(\cdot | \cdot)$ denotes the rank association measure Somers’ $D$.

- *In other words*, a median slope is a linear effect of $X$ on $Y$, large enough to explain the observed association.

- If $X$ is binary with values 0 and 1, then the Theil–Sen median slope is the Hodges–Lehmann median difference between the subpopulations in which $X = 1$ and $X = 0$.

- *In other words*, the Hodges–Lehmann median difference is the median pairwise difference between two $Y$–values, sampled at random from the two subpopulations.

- Note that the median difference is *not* always the difference between the two subpopulation medians!
What is a Hodges–Lehmann median difference?

- A **Theil–Sen median slope** of $Y$ with respect to $X$ is a solution in $\beta$ to the equation $D(Y - \beta X|X) = 0$, where $D(\cdot|\cdot)$ denotes the rank association measure Somers’ $D$.

- **In other words**, a median slope is a linear effect of $X$ on $Y$, large enough to explain the observed association.

- If $X$ is binary with values 0 and 1, then the Theil–Sen median slope is the **Hodges–Lehmann median difference** between the subpopulations in which $X = 1$ and $X = 0$.

- **In other words**, the Hodges–Lehmann median difference is the median pairwise difference between two $Y$–values, sampled at random from the two subpopulations.

- Note that the median difference is *not* always the difference between the two subpopulation medians!
What is a Hodges–Lehmann median difference?

- A **Theil–Sen median slope** of $Y$ with respect to $X$ is a solution in $\beta$ to the equation $D(Y - \beta X|X) = 0$, where $D(\cdot|\cdot)$ denotes the rank association measure Somers’ $D$.

- *In other words*, a median slope is a linear effect of $X$ on $Y$, large enough to explain the observed association.

- If $X$ is binary with values 0 and 1, then the Theil–Sen median slope is the **Hodges–Lehmann median difference** between the subpopulations in which $X = 1$ and $X = 0$.

- *In other words*, the Hodges–Lehmann median difference is the median pairwise difference between two $Y$–values, sampled at random from the two subpopulations.

- Note that the median difference is *not* always the difference between the two subpopulation medians!
What is a Hodges–Lehmann median difference?

- A **Theil–Sen median slope** of $Y$ with respect to $X$ is a solution in $\beta$ to the equation $D(Y - \beta X | X) = 0$, where $D(\cdot | \cdot)$ denotes the rank association measure Somers’ $D$.

- *In other words*, a median slope is a linear effect of $X$ on $Y$, large enough to explain the observed association.

- If $X$ is binary with values 0 and 1, then the Theil–Sen median slope is the **Hodges–Lehmann median difference** between the subpopulations in which $X = 1$ and $X = 0$.

- *In other words*, the Hodges–Lehmann median difference is the median pairwise difference between two $Y$–values, sampled at random from the two subpopulations.

- Note that the median difference is *not* always the difference between the two subpopulation medians!
What is a Hodges–Lehmann median difference?

- A Theil–Sen median slope of $Y$ with respect to $X$ is a solution in $\beta$ to the equation $D(Y - \beta X | X) = 0$, where $D(\cdot | \cdot)$ denotes the rank association measure Somers’ $D$.

- In other words, a median slope is a linear effect of $X$ on $Y$, large enough to explain the observed association.

- If $X$ is binary with values 0 and 1, then the Theil–Sen median slope is the Hodges–Lehmann median difference between the subpopulations in which $X = 1$ and $X = 0$.

- In other words, the Hodges–Lehmann median difference is the median pairwise difference between two $Y$–values, sampled at random from the two subpopulations.

- Note that the median difference is not always the difference between the two subpopulation medians!
The Lehmann confidence interval formula

- The conventional confidence interval formula for the median difference (Lehmann, 1963)[1] was implemented in Stata by Wang (1999)[4].

- It assumes that the two subpopulation distributions are different only in location.

- This assumption implies that the median difference is the difference between the two medians.

- However, it also implies that the subpopulations are equally variable.

- The Lehmann formula is therefore robust to non-Normality at the price of being non-robust to unequal variability. (Which often causes even more problems.)
The conventional confidence interval formula for the median difference (Lehmann, 1963)[1] was implemented in Stata by Wang (1999)[4].

It assumes that the two subpopulation distributions are different only in location.

This assumption implies that the median difference is the difference between the two medians.

However, it also implies that the subpopulations are equally variable.

The Lehmann formula is therefore robust to non–Normality at the price of being non–robust to unequal variability. (Which often causes even more problems.)
The Lehmann confidence interval formula

▶ The conventional confidence interval formula for the median difference (Lehmann, 1963)[1] was implemented in Stata by Wang (1999)[4].
▶ It assumes that the two subpopulation distributions are different only in location.
▶ This assumption implies that the median difference is the difference between the two medians.
▶ However, it also implies that the subpopulations are equally variable.
▶ The Lehmann formula is therefore robust to non–Normality at the price of being non–robust to unequal variability. (Which often causes even more problems.)
The Lehmann confidence interval formula

- The conventional confidence interval formula for the median difference (Lehmann, 1963)[1] was implemented in Stata by Wang (1999)[4].
- It assumes that the two subpopulation distributions are different only in location.
- This assumption implies that the median difference is the difference between the two medians.
- However, it also implies that the subpopulations are equally variable.
- The Lehmann formula is therefore robust to non-Normality at the price of being non-robust to unequal variability. (Which often causes even more problems.)
The Lehmann confidence interval formula

- The conventional confidence interval formula for the median difference (Lehmann, 1963)[1] was implemented in Stata by Wang (1999)[4].
- It assumes that the two subpopulation distributions are different only in location.
- This assumption implies that the median difference is the difference between the two medians.
- However, it also implies that the subpopulations are equally variable.
- The Lehmann formula is therefore robust to non-Normality at the price of being non-robust to unequal variability. (Which often causes even more problems.)
The conventional confidence interval formula for the median difference (Lehmann, 1963) was implemented in Stata by Wang (1999).

It assumes that the two subpopulation distributions are different only in location.

This assumption implies that the median difference is the difference between the two medians.

However, it also implies that the subpopulations are equally variable.

The Lehmann formula is therefore robust to non-Normality at the price of being non-robust to unequal variability. (Which often causes even more problems.)
The `cendif` confidence interval formula

- An alternative confidence interval formula for the median difference (Newson, 2006) is used by the `cendif` module of the SSC package `somersd`.
- It is derived by inverting a delta–jackknife confidence interval formula for Somers’ $D$.
- It should therefore still work if the two subpopulation distributions differ in ways other than location.
- In particular, it should still work if the two subpopulations are unequally variable.
- The `cendif` formula therefore contrasts to the Lehmann formula as the unequal–variance $t$–test contrasts to the equal–variance $t$–test.
The `cendif` confidence interval formula

- An alternative confidence interval formula for the median difference (Newson, 2006) is used by the `cendif` module of the `somersd` package.
- It is derived by inverting a delta–jackknife confidence interval formula for Somers’ $D$.
- It should therefore still work if the two subpopulation distributions differ in ways other than location.
- In particular, it should still work if the two subpopulations are unequally variable.
- The `cendif` formula therefore contrasts to the Lehmann formula as the unequal–variance $t$–test contrasts to the equal–variance $t$–test.
The `cendif` confidence interval formula

- An alternative confidence interval formula for the median difference (Newson, 2006)[3] is used by the `cendif` module of the SSC package `somersd`.
- It is derived by inverting a delta–jackknife confidence interval formula for Somers’ $D$.
- It should therefore still work if the two subpopulation distributions differ in ways other than location.
- In particular, it should still work if the two subpopulations are unequally variable.
- The `cendif` formula therefore contrasts to the Lehmann formula as the unequal–variance $t$–test contrasts to the equal–variance $t$–test.
The \texttt{cendif} confidence interval formula

- An alternative confidence interval formula for the median difference (Newson, 2006)\cite{Newson2006} is used by the \texttt{cendif} module of the SSC package \texttt{somersd}.
- It is derived by inverting a delta–jackknife confidence interval formula for Somers’ $D$.
- It should therefore still work if the two subpopulation distributions differ in ways other than location.
- In particular, it should still work if the two subpopulations are unequally variable.
- The \texttt{cendif} formula therefore contrasts to the Lehmann formula as the unequal–variance $t$–test contrasts to the equal–variance $t$–test.
The `cendif` confidence interval formula

- An alternative confidence interval formula for the median difference (Newson, 2006)\cite{newson2006} is used by the `cendif` module of the SSC package `somersd`.
- It is derived by inverting a delta–jackknife confidence interval formula for Somers’ $D$.
- It should therefore still work if the two subpopulation distributions differ in ways other than location.
- In particular, it should still work if the two subpopulations are unequally variable.
- The `cendif` formula therefore contrasts to the Lehmann formula as the unequal–variance $t$–test contrasts to the equal–variance $t$–test.
The `cendif` confidence interval formula

- An alternative confidence interval formula for the median difference (Newson, 2006)\cite{Newson2006} is used by the `cendif` module of the SSC package `somersd`.
- It is derived by inverting a delta–jackknife confidence interval formula for Somers’ $D$.
- It should therefore still work if the two subpopulation distributions differ in ways other than location.
- In particular, it should still work if the two subpopulations are unequally variable.
- The `cendif` formula therefore contrasts to the Lehmann formula as the unequal–variance $t$–test contrasts to the equal–variance $t$–test.
Comparing the two $t$–tests: Existing results

- The Satterthwaite method had the advertised coverage probability.
- The equal–variance $t$–test produced oversized (undersized) confidence intervals if the smaller sample is sampled from the less variable (more variable) subpopulation.
- However, the equal–variance $t$–test had the advertised coverage probability, if either the subsample numbers or the subpopulation variances were equal.
- Under the latter conditions, the equal–variance $t$–test produced smaller confidence intervals with the same coverage probability.
- The authors therefore recommended the unequal–variance method as the “default”, and the equal–variance method for the “special occasion” of unequal sample numbers and prior knowledge of equal variability.
- They advised against the “traditional” practice of testing equality of variances before choosing a $t$–test!
Comparing the two $t$–tests: Existing results

- The Satterthwaite method had the advertised coverage probability.
- The equal–variance $t$–test produced oversized (undersized) confidence intervals if the smaller sample is sampled from the less variable (more variable) subpopulation.
- However, the equal–variance $t$–test had the advertised coverage probability, if either the subsample numbers or the subpopulation variances were equal.
- Under the latter conditions, the equal–variance $t$–test produced smaller confidence intervals with the same coverage probability.
- The authors therefore recommended the unequal–variance method as the “default”, and the equal–variance method for the “special occasion” of unequal sample numbers and prior knowledge of equal variability.
- They advised against the “traditional” practice of testing equality of variances before choosing a $t$–test!
Comparing the two \( t \)–tests: Existing results

- The Satterthwaite method had the advertised coverage probability.
- The equal–variance \( t \)–test produced oversized (undersized) confidence intervals if the smaller sample is sampled from the less variable (more variable) subpopulation.
- However, the equal–variance \( t \)–test had the advertised coverage probability, if either the subsample numbers or the subpopulation variances were equal.
- Under the latter conditions, the equal–variance \( t \)–test produced smaller confidence intervals with the same coverage probability.
- The authors therefore recommended the unequal–variance method as the “default”, and the equal–variance method for the “special occasion” of unequal sample numbers and prior knowledge of equal variability.
- They advised against the “traditional” practice of testing equality of variances before choosing a \( t \)–test!
Comparing the two $t$–tests: Existing results

- The Satterthwaite method had the advertised coverage probability.
- The equal–variance $t$–test produced oversized (undersized) confidence intervals if the smaller sample is sampled from the less variable (more variable) subpopulation.
- However, the equal–variance $t$–test had the advertised coverage probability, if either the subsample numbers or the subpopulation variances were equal.
- Under the latter conditions, the equal–variance $t$–test produced smaller confidence intervals with the same coverage probability.
- The authors therefore recommended the unequal–variance method as the “default”, and the equal–variance method for the “special occasion” of unequal sample numbers and prior knowledge of equal variability.
- They advised against the “traditional” practice of testing equality of variances before choosing a $t$–test!
Comparing the two $t$–tests: Existing results


► The Satterthwaite method had the advertised coverage probability.

► The equal–variance $t$–test produced oversized (undersized) confidence intervals if the smaller sample is sampled from the less variable (more variable) subpopulation.

► *However*, the equal–variance $t$–test had the advertised coverage probability, if *either* the subsample numbers *or* the subpopulation variances were equal.

► Under the *latter* conditions, the equal–variance $t$–test produced smaller confidence intervals with the same coverage probability.

► The authors therefore recommended the unequal–variance method as the “default”, and the equal–variance method for the “special occasion” of unequal sample numbers and prior knowledge of equal variability.

► They advised *against* the “traditional” practice of testing equality of variances *before* choosing a $t$–test!
Comparing the two $t$–tests: Existing results

- The Satterthwaite method had the advertised coverage probability.
- The equal–variance $t$–test produced oversized (undersized) confidence intervals if the smaller sample is sampled from the less variable (more variable) subpopulation.
- *However*, the equal–variance $t$–test had the advertised coverage probability, if *either* the subsample numbers *or* the subpopulation variances were equal.
- Under the *latter* conditions, the equal–variance $t$–test produced smaller confidence intervals with the same coverage probability.
- The authors therefore recommended the unequal–variance method as the “default”, and the equal–variance method for the “special occasion” of unequal sample numbers and *prior* knowledge of equal variability.
- They advised *against* the “traditional” practice of testing equality of variances *before* choosing a $t$–test!
Comparing the two $t$–tests: Existing results

- The Satterthwaite method had the advertised coverage probability.
- The equal–variance $t$–test produced oversized (undersized) confidence intervals if the smaller sample is sampled from the less variable (more variable) subpopulation.
- However, the equal–variance $t$–test had the advertised coverage probability, if either the subsample numbers or the subpopulation variances were equal.
- Under the latter conditions, the equal–variance $t$–test produced smaller confidence intervals with the same coverage probability.
- The authors therefore recommended the unequal–variance method as the “default”, and the equal–variance method for the “special occasion” of unequal sample numbers and prior knowledge of equal variability.
- They advised against the “traditional” practice of testing equality of variances before choosing a $t$–test!
Comparing the two $t$–tests: Existing results


- The Satterthwaite method had the advertised coverage probability.

- The equal–variance $t$–test produced oversized (undersized) confidence intervals if the smaller sample is sampled from the less variable (more variable) subpopulation.

- $\textit{However}$, the equal–variance $t$–test had the advertised coverage probability, if either the subsample numbers or the subpopulation variances were equal.

- Under the $\textit{latter}$ conditions, the equal–variance $t$–test produced smaller confidence intervals with the same coverage probability.

- The authors therefore recommended the unequal–variance method as the “default”, and the equal–variance method for the “special occasion” of unequal sample numbers and prior knowledge of equal variability.

- They advised $\textit{against}$ the “traditional” practice of testing equality of variances before choosing a $t$–test!
Simulation study: Aims

- A simulation study, modelled on the Moser–Stevens study[2], was designed to test \textit{cendif} to destruction in a wide range of scenarios.
- The \textit{cendif} method was compared with 3 other methods (the Lehmann method and the two \textit{t}–tests) for calculating confidence intervals for median differences.
- In each scenario, coverage probabilities were estimated, together with median confidence interval width ratios.
- 10000 replicate sample pairs were simulated for each scenario.
- In this presentation, we focus on comparing coverage probabilities between the Lehmann and \textit{cendif} methods.
Simulation study: Aims

▶ A simulation study, modelled on the Moser–Stevens study[2], was designed to test \texttt{cendif} to destruction in a wide range of scenarios.

▶ The \texttt{cendif} method was compared with 3 other methods (the Lehmann method and the two \textit{t}–tests) for calculating confidence intervals for median differences.

▶ In each scenario, coverage probabilities were estimated, together with median confidence interval width ratios.

▶ 10000 replicate sample pairs were simulated for each scenario.

▶ In this presentation, we focus on comparing coverage probabilities between the Lehmann and \texttt{cendif} methods.
Simulation study: Aims

- A simulation study, modelled on the Moser–Stevens study[2], was designed to test \texttt{cendif} to destruction in a wide range of scenarios.

- The \texttt{cendif} method was compared with 3 other methods (the Lehmann method and the two $t$–tests) for calculating confidence intervals for median differences.

- In each scenario, coverage probabilities were estimated, together with median confidence interval width ratios.

- 10000 replicate sample pairs were simulated for each scenario.

- In this presentation, we focus on comparing coverage probabilities between the Lehmann and \texttt{cendif} methods.
Simulation study: Aims

- A simulation study, modelled on the Moser–Stevens study[2], was designed to test cendif to destruction in a wide range of scenarios.

- The cendif method was compared with 3 other methods (the Lehmann method and the two $t$-tests) for calculating confidence intervals for median differences.

- In each scenario, coverage probabilities were estimated, together with median confidence interval width ratios.

- 10000 replicate sample pairs were simulated for each scenario.

- In this presentation, we focus on comparing coverage probabilities between the Lehmann and cendif methods.
Simulation study: Aims

- A simulation study, modelled on the Moser–Stevens study[2], was designed to test `cendif` to destruction in a wide range of scenarios.
- The `cendif` method was compared with 3 other methods (the Lehmann method and the two \(t\)-tests) for calculating confidence intervals for median differences.
- In each scenario, coverage probabilities were estimated, together with median confidence interval width ratios.
- 10,000 replicate sample pairs were simulated for each scenario.
- In this presentation, we focus on comparing coverage probabilities between the Lehmann and `cendif` methods.
Simulation study: Aims

▶ A simulation study, modelled on the Moser–Stevens study[2], was designed to test cendif to destruction in a wide range of scenarios.

▶ The cendif method was compared with 3 other methods (the Lehmann method and the two t–tests) for calculating confidence intervals for median differences.

▶ In each scenario, coverage probabilities were estimated, together with median confidence interval width ratios.

▶ 10000 replicate sample pairs were simulated for each scenario.

▶ In this presentation, we focus on comparing coverage probabilities between the Lehmann and cendif methods.
Simulation study: Scenarios

- Pairs of subpopulation distributions were selected from 2 families: the “t–test friendly” Normal family and the outlier–prone, “t–test unfriendly” Cauchy family.
- Both families are symmetric, and parameterized by a median $\mu$ (set to zero) and a scale parameter $\sigma$ (measuring variability).
- Subsample numbers were all 10 possible pairs $N_1 \leq N_2$ from the set \{5, 10, 20, 40\}.
- Variability scale ratios $\sigma_1/\sigma_2$ between the populations of the smaller and larger samples were from the symmetrical set of 9 values \{1/4, 1/3, 1/2, 2/3, 1, 3/2, 2, 3, 4\}.
- These 180 scenarios (90 for each distributional family) were chosen to include “best” and “worst” cases for all 4 statistical methods.
Simulation study: Scenarios

- Pairs of subpopulation distributions were selected from 2 families: the “t–test friendly” Normal family and the outlier–prone, “t–test unfriendly” Cauchy family.
- Both families are symmetric, and parameterized by a median $\mu$ (set to zero) and a scale parameter $\sigma$ (measuring variability).
- Subsample numbers were all 10 possible pairs $N_1 \leq N_2$ from the set $\{5, 10, 20, 40\}$.
- Variability scale ratios $\sigma_1/\sigma_2$ between the populations of the smaller and larger samples were from the symmetrical set of 9 values $\{1/4, 1/3, 1/2, 2/3, 1, 3/2, 2, 3, 4\}$.
- These 180 scenarios (90 for each distributional family) were chosen to include “best” and “worst” cases for all 4 statistical methods.
Simulation study: Scenarios

- Pairs of subpopulation distributions were selected from 2 families: the “t–test friendly” Normal family and the outlier–prone, “t–test unfriendly” Cauchy family.
- Both families are symmetric, and parameterized by a median $\mu$ (set to zero) and a scale parameter $\sigma$ (measuring variability).
- Subsample numbers were all 10 possible pairs $N_1 \leq N_2$ from the set \{5, 10, 20, 40\}.
- Variability scale ratios $\sigma_1/\sigma_2$ between the populations of the smaller and larger samples were from the symmetrical set of 9 values \{1/4, 1/3, 1/2, 2/3, 1, 3/2, 2, 3, 4\}.
- These 180 scenarios (90 for each distributional family) were chosen to include “best” and “worst” cases for all 4 statistical methods.
Simulation study: Scenarios

- Pairs of subpopulation distributions were selected from 2 families: the “$t$–test friendly” Normal family and the outlier–prone, “$t$–test unfriendly” Cauchy family.
- Both families are symmetric, and parameterized by a median $\mu$ (set to zero) and a scale parameter $\sigma$ (measuring variability).
- Subsample numbers were all 10 possible pairs $N_1 \leq N_2$ from the set $\{5, 10, 20, 40\}$.
- Variability scale ratios $\sigma_1/\sigma_2$ between the populations of the smaller and larger samples were from the symmetrical set of 9 values $\{1/4, 1/3, 1/2, 2/3, 1, 3/2, 2, 3, 4\}$.
- These 180 scenarios (90 for each distributional family) were chosen to include “best” and “worst” cases for all 4 statistical methods.
Simulation study: Scenarios

- Pairs of subpopulation distributions were selected from 2 families: the “t–test friendly” Normal family and the outlier–prone, “t–test unfriendly” Cauchy family.
- Both families are symmetric, and parameterized by a median $\mu$ (set to zero) and a scale parameter $\sigma$ (measuring variability).
- Subsample numbers were all 10 possible pairs $N_1 \leq N_2$ from the set $\{5, 10, 20, 40\}$.
- Variability scale ratios $\sigma_1/\sigma_2$ between the populations of the smaller and larger samples were from the symmetrical set of 9 values $\{1/4, 1/3, 1/2, 2/3, 1, 3/2, 2, 3, 4\}$.
- These 180 scenarios (90 for each distributional family) were chosen to include “best” and “worst” cases for all 4 statistical methods.
Simulation study: Scenarios

- Pairs of subpopulation distributions were selected from 2 families: the “*t*-test friendly” Normal family and the outlier-prone, “*t*-test unfriendly” Cauchy family.
- Both families are symmetric, and parameterized by a median $\mu$ (set to zero) and a scale parameter $\sigma$ (measuring variability).
- Subsample numbers were all 10 possible pairs $N_1 \leq N_2$ from the set $\{5, 10, 20, 40\}$.
- Variability scale ratios $\sigma_1/\sigma_2$ between the populations of the smaller and larger samples were from the symmetrical set of 9 values $\{1/4, 1/3, 1/2, 2/3, 1, 3/2, 2, 3, 4\}$.
- These 180 scenarios (90 for each distributional family) were chosen to include “best” and “worst” cases for all 4 statistical methods.
Normal coverage probabilities for the Gosset and \textit{cendif} methods

<table>
<thead>
<tr>
<th>First/Second population scale ratio</th>
<th>Gosset coverage</th>
<th>cendif coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>5, 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5, 10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5, 20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5, 40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10, 10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10, 20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10, 40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20, 20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20, 40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40, 40</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Coverage probability under Normal distribution

The equal–variance \textit{t}–test produces oversized (undersized) confidence intervals if the smaller sample is from the less (more) variable population.

Robust confidence intervals for Hodges-Lehmann median differences
Normal coverage probabilities for the Lehmann and `cendif` methods

<table>
<thead>
<tr>
<th>First/second population scale ratio</th>
<th>Coverage probability under Normal distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>5, 5</td>
<td><img src="image" alt="Graph" /></td>
</tr>
<tr>
<td>10, 20</td>
<td><img src="image" alt="Graph" /></td>
</tr>
<tr>
<td>20, 20</td>
<td><img src="image" alt="Graph" /></td>
</tr>
</tbody>
</table>

Graphs by First sample number and Second sample number

Under most (but not all) scenarios, the `cendif` coverage probability is closer to the advertised value of 0.95.
For both rank methods, the Cauchy coverage probabilities are similar to the Normal coverage probabilities. However . . .
Lehmann versus cendif: Patterns of relative advantage

... the relative advantage between the two rank methods varies between scenarios.

The subsample size pairs $N_1 \leq N_2$ can be classified into 3 “fuzzy patterns”, which blend into each other gradually.

These 3 patterns can be named “$N_1 = N_2$”, “$N_1 < N_2$”, and “$N_1 \ll N_2$”.

We will illustrate this remark by focussing on a “typical” example of each pattern.
Lehmann versus cendif: Patterns of relative advantage

- ... the relative advantage between the two rank methods varies between scenarios.

- The subsample size pairs $N_1 \leq N_2$ can be classified into 3 “fuzzy patterns”, which blend into each other gradually.

- These 3 patterns can be named “$N_1 = N_2$”, “$N_1 < N_2$”, and “$N_1 \ll N_2$”.

- We will illustrate this remark by focussing on a “typical” example of each pattern.
Lehmann versus cendif: Patterns of relative advantage

- The relative advantage between the two rank methods varies between scenarios.
- The subsample size pairs $N_1 \leq N_2$ can be classified into 3 “fuzzy patterns”, which blend into each other gradually.
- These 3 patterns can be named “$N_1 = N_2$”, “$N_1 < N_2$”, and “$N_1 \ll N_2$”.
- We will illustrate this remark by focusing on a “typical” example of each pattern.
... the relative advantage between the two rank methods varies between scenarios.

The subsample size pairs $N_1 \leq N_2$ can be classified into 3 “fuzzy patterns”, which blend into each other gradually.

These 3 patterns can be named “$N_1 = N_2$”, “$N_1 < N_2$”, and “$N_1 \ll N_2$”.

We will illustrate this remark by focussing on a “typical” example of each pattern.
$N_1 = N_2$: Both methods are reasonable

- Median differences between 2 Normal samples of 40 are estimated.
- Both methods have coverage probabilities close to the advertised level of 0.95.
- However, the Lehmann method produces slightly undersized confidence intervals under very unequal variability.
\( N_1 = N_2: \) Both methods are reasonable

- Median differences between 2 Normal samples of 40 are estimated.
- Both methods have coverage probabilities close to the advertised level of 0.95.
- However, the Lehmann method produces slightly undersized confidence intervals under very unequal variability.
$N_1 = N_2$: Both methods are reasonable

- Median differences between 2 Normal samples of 40 are estimated.

- Both methods have coverage probabilities close to the advertised level of 0.95.

- However, the Lehmann method produces slightly undersized confidence intervals under very unequal variability.
$N_1 = N_2$: Both methods are reasonable

- Median differences between 2 Normal samples of 40 are estimated.
- Both methods have coverage probabilities close to the advertised level of 0.95.
- However, the Lehmann method produces slightly undersized confidence intervals under very unequal variability.
$N_1 < N_2$: *cendif* is robust

- The first sample number here is half the second.
- The *cendif* method has coverage probabilities close to the advertised level of 0.95 under all variability ratios.
- The Lehmann method produces oversized (undersized) confidence intervals if the smaller sample is from the less (more) variable population. (Like the equal–variance t–test.)
$N_1 < N_2$: \texttt{cendif} is robust

- The first sample number here is half the second.
- The \texttt{cendif} method has coverage probabilities close to the advertised level of 0.95 under all variability ratios.
- The Lehmann method produces oversized (undersized) confidence intervals if the smaller sample is from the less (more) variable population. (Like the equal-variance $t$-test.)

![Graph showing coverage probabilities under Normal distribution](image)
$N_1 < N_2$: \textit{cendif} is robust

- The first sample number here is half the second.
- The \textit{cendif} method has coverage probabilities close to the advertised level of 0.95 under all variability ratios.
- The Lehmann method produces oversized (undersized) confidence intervals if the smaller sample is from the less (more) variable population. (Like the equal–variance $t$–test.)

Robust confidence intervals for Hodges-Lehmann median differences
$N_1 < N_2$: **cendif** is robust

- The first sample number here is half the second.
- The **cendif** method has coverage probabilities close to the advertised level of 0.95 under all variability ratios.
- The Lehmann method produces oversized (undersized) confidence intervals if the smaller sample is from the less (more) variable population. (Like the equal–variance $t$–test.)
$N_1 \ll N_2$: \texttt{cendif} is tested to destruction

- The \texttt{cendif} confidence interval is now undersized under most variability ratios.
- The Lehmann method still produces oversized (undersized) confidence intervals if the smaller sample is from the less (more) variable population.
- \textit{However}, the Lehmann coverage is at least correct under equal variability!
$N_1 \ll N_2$: `cendif` is tested to destruction

- The `cendif` confidence interval is now undersized under most variability ratios.

- The Lehmann method still produces oversized (undersized) confidence intervals if the smaller sample is from the less (more) variable population.

- However, the Lehmann coverage is at least correct under equal variability!

---

Robust confidence intervals for Hodges-Lehmann median differences

Graphs by First sample number and Second sample number
$N_1 \ll N_2$: \textbf{cendif} is tested to destruction

- The \textbf{cendif} confidence interval is now undersized under most variability ratios.

- The Lehmann method still produces oversized (undersized) confidence intervals if the smaller sample is from the less (more) variable population.

- \textit{However,} the Lehmann coverage is at least correct under equal variability!
$N_1 \ll N_2$: `cendif' is tested to destruction

- The `cendif' confidence interval is now undersized under most variability ratios.
- The Lehmann method still produces oversized (undersized) confidence intervals if the smaller sample is from the less (more) variable population.
- *However*, the Lehmann coverage is at least correct under equal variability!

![Graph showing Lehmann and cendif coverage probabilities](image-url)
Lehmann versus cendif: Summary of results

- If $N_1 = N_2$, then both methods (especially cendif) produce coverage probabilities close to the advertised level.

- If $N_1 < N_2$ (and $N_1$ is not too small), then the Lehmann method produces oversized (undersized) confidence intervals if the smaller sample is from the less (more) variable population, and the cendif method is more robust.

- However, if $N_1 \ll N_2$ (and $N_1$ is very small), then the cendif method produces undersized confidence intervals, and the Lehmann method is more correct under equal variability.

- Therefore, cendif is robust to unequal variability, at the price of being less robust to the possibility that the smaller sample (but not the larger one) is very small.
Lehmann versus cendif: Summary of results

▶ If $N_1 = N_2$, then both methods (especially cendif) produce coverage probabilities close to the advertized level.

▶ If $N_1 < N_2$ (and $N_1$ is not too small), then the Lehmann method produces oversized (undersized) confidence intervals if the smaller sample is from the less (more) variable population, and the cendif method is more robust.

▶ However, if $N_1 \ll N_2$ (and $N_1$ is very small), then the cendif method produces undersized confidence intervals, and the Lehmann method is more correct under equal variability.

▶ Therefore, cendif is robust to unequal variability, at the price of being less robust to the possibility that the smaller sample (but not the larger one) is very small.
Lehmann versus cendif: Summary of results

▶ If $N_1 = N_2$, then both methods (especially cendif) produce coverage probabilities close to the advertised level.

▶ If $N_1 < N_2$ (and $N_1$ is not too small), then the Lehmann method produces oversized (undersized) confidence intervals if the smaller sample is from the less (more) variable population, and the cendif method is more robust.

▶ However, if $N_1 \ll N_2$ (and $N_1$ is very small), then the cendif method produces undersized confidence intervals, and the Lehmann method is more correct under equal variability.

▶ Therefore, cendif is robust to unequal variability, at the price of being less robust to the possibility that the smaller sample (but not the larger one) is very small.
Lehmann versus \texttt{cendif}: Summary of results

- If $N_1 = N_2$, then both methods (especially \texttt{cendif}) produce coverage probabilities close to the advertised level.

- If $N_1 < N_2$ (and $N_1$ is not too small), then the Lehmann method produces oversized (undersized) confidence intervals if the smaller sample is from the less (more) variable population, and the \texttt{cendif} method is more robust.

- \textit{However}, if $N_1 \ll N_2$ (and $N_1$ is very small), then the \texttt{cendif} method produces undersized confidence intervals, and the Lehmann method is more correct \textit{under equal variability}.

- \textit{Therefore}, \texttt{cendif} is robust to unequal variability, at the price of being less robust to the possibility that the smaller sample (but not the larger one) is very small.
Lehmann *versus* `cendif`: Summary of results

- If $N_1 = N_2$, then both methods (especially `cendif`) produce coverage probabilities close to the advertised level.
- If $N_1 < N_2$ (and $N_1$ is not too small), then the Lehmann method produces oversized (undersized) confidence intervals if the smaller sample is from the less (more) variable population, and the `cendif` method is more robust.
- *However*, if $N_1 \ll N_2$ (and $N_1$ is very small), then the `cendif` method produces undersized confidence intervals, and the Lehmann method is more correct under equal variability.
- *Therefore*, `cendif` is robust to unequal variability, at the price of being less robust to the possibility that the smaller sample (but not the larger one) is very small.
The Lehmann and `cendif` methods are both based on Central Limit Theorems, applied to Somers’ $D(Y|X)$ for a binary $X$ and a continuous $Y$.

However, the `cendif` method estimates the variance from the joint sample distribution of $X$ and $Y$, using jackknife methods.

By contrast, the Lehmann method calculates the variance from the marginal sample distributions of $X$ and $Y$, using permutation methods.

Therefore, the Lehmann method (like the equal–variance $t$–test) estimates the population variability of the smaller sample using the sample variability of the larger sample.

By contrast, the `cendif` method (like the unequal–variance $t$–test) estimates the population variability of the smaller sample using the sample variability of the smaller sample.
The Lehmann and cendif methods are both based on Central Limit Theorems, applied to Somers’ $D(Y|X)$ for a binary $X$ and a continuous $Y$.

However, the cendif method estimates the variance from the joint sample distribution of $X$ and $Y$, using jackknife methods.

By contrast, the Lehmann method calculates the variance from the marginal sample distributions of $X$ and $Y$, using permutation methods.

Therefore, the Lehmann method (like the equal–variance $t$–test) estimates the population variability of the smaller sample using the sample variability of the larger sample.

By contrast, the cendif method (like the unequal–variance $t$–test) estimates the population variability of the smaller sample using the sample variability of the smaller sample.
Lehmann *versus* cendif: General principles

- The Lehmann and cendif methods are both based on Central Limit Theorems, applied to Somers’ $D(Y|X)$ for a binary $X$ and a continuous $Y$.

- *However*, the cendif method *estimates* the variance from the *joint* sample distribution of $X$ and $Y$, using jackknife methods.

- By contrast, the Lehmann method *calculates* the variance from the *marginal* sample distributions of $X$ and $Y$, using permutation methods.

- *Therefore*, the Lehmann method (like the equal–variance $t$–test) estimates the *population* variability of the smaller sample using the *sample* variability of the *larger* sample.

- By contrast, the cendif method (like the unequal–variance $t$–test) estimates the population variability of the smaller sample using the sample variability of the *smaller* sample.
Lehmann \textit{versus} \texttt{cendif}: General principles

- The Lehmann and \texttt{cendif} methods are both based on Central Limit Theorems, applied to Somers’ $D(Y|X)$ for a binary $X$ and a continuous $Y$.

- \textit{However}, the \texttt{cendif} method \textit{estimates} the variance from the \textit{joint} sample distribution of $X$ and $Y$, using jackknife methods.

- By contrast, the Lehmann method \textit{calculates} the variance from the \textit{marginal} sample distributions of $X$ and $Y$, using permutation methods.

- \textit{Therefore}, the Lehmann method (like the equal–variance $t$–test) estimates the \textit{population} variability of the smaller sample using the \textit{sample} variability of the larger sample.

- By contrast, the \texttt{cendif} method (like the unequal–variance $t$–test) estimates the population variability of the smaller sample using the sample variability of the smaller sample.
Lehmann versus cendif: General principles

- The Lehmann and cendif methods are both based on Central Limit Theorems, applied to Somers’ $D(Y|X)$ for a binary $X$ and a continuous $Y$.

- **However**, the cendif method *estimates* the variance from the *joint* sample distribution of $X$ and $Y$, using jackknife methods.

- By contrast, the Lehmann method *calculates* the variance from the *marginal* sample distributions of $X$ and $Y$, using permutation methods.

- **Therefore**, the Lehmann method (like the equal–variance $t$–test) estimates the *population* variability of the smaller sample using the *sample* variability of the *larger* sample.

- By contrast, the cendif method (like the unequal–variance $t$–test) estimates the population variability of the smaller sample using the sample variability of the *smaller* sample.
Lehmann versus cendif: General principles

- The Lehmann and cendif methods are both based on Central Limit Theorems, applied to Somers’ $D(Y|X)$ for a binary $X$ and a continuous $Y$.

- **However**, the cendif method *estimates* the variance from the *joint* sample distribution of $X$ and $Y$, using jackknife methods.

- By contrast, the Lehmann method *calculates* the variance from the *marginal* sample distributions of $X$ and $Y$, using permutation methods.

- **Therefore**, the Lehmann method (like the equal–variance $t$–test) estimates the population variability of the smaller sample using the sample variability of the *larger* sample.

- By contrast, the cendif method (like the unequal–variance $t$–test) estimates the population variability of the smaller sample using the sample variability of the *smaller* sample.
Lehmann versus `cendif`: Interpretation of results

- If $N_1 = N_2$, then there is no larger or smaller sample – and both methods work (especially `cendif`).
- If $N_1 < N_2$ (and $N_1$ is not too small), then the population variability of the smaller sample is best estimated using the sample variability of the smaller sample – favoring `cendif`.
- If $N_1 \ll N_2$ (and $N_1$ is very small), and we have prior reason to expect “similar” variability, then the population variability of the smaller sample is best estimated using the sample variability of the larger sample – favoring the Lehmann method.
- This seems to suggest a policy of regarding `cendif` as the default and the Lehmann formula as the “special case”, similar to the Moser–Stevens[2] policy regarding the two $t$–tests.
Lehmann versus cendif: Interpretation of results

- If $N_1 = N_2$, then there is no larger or smaller sample – and both methods work (especially cendif).
- If $N_1 < N_2$ (and $N_1$ is not too small), then the population variability of the smaller sample is best estimated using the sample variability of the smaller sample – favoring cendif.
- If $N_1 \ll N_2$ (and $N_1$ is very small), and we have prior reason to expect “similar” variability, then the population variability of the smaller sample is best estimated using the sample variability of the larger sample – favoring the Lehmann method.
- This seems to suggest a policy of regarding cendif as the default and the Lehmann formula as the “special case”, similar to the Moser–Stevens[2] policy regarding the two $t$–tests.
Lehmann *versus cendif*: Interpretation of results

- If $N_1 = N_2$, then there is no larger or smaller sample – and both methods work (especially *cendif*).

- If $N_1 < N_2$ (and $N_1$ is not too small), then the population variability of the smaller sample is best estimated using the sample variability of the smaller sample – favoring *cendif*.

- If $N_1 \ll N_2$ (and $N_1$ is very small), and we have prior reason to expect “similar” variability, then the population variability of the smaller sample is best estimated using the sample variability of the larger sample – favoring the Lehmann method.

- This seems to suggest a policy of regarding *cendif* as the default and the Lehmann formula as the “special case”, similar to the Moser–Stevens[2] policy regarding the two $t$–tests.
Lehmann versus \texttt{cendif}: Interpretation of results

- If $N_1 = N_2$, then there is no larger or smaller sample – and both methods work (especially \texttt{cendif}).
- If $N_1 < N_2$ (and $N_1$ is not too small), then the population variability of the smaller sample is best estimated using the sample variability of the smaller sample – favoring \texttt{cendif}.
- If $N_1 \ll N_2$ (and $N_1$ is very small), and we have prior reason to expect “similar” variability, then the population variability of the smaller sample is best estimated using the sample variability of the larger sample – favoring the Lehmann method.
- This seems to suggest a policy of regarding \texttt{cendif} as the default and the Lehmann formula as the “special case”, similar to the Moser–Stevens\cite{2} policy regarding the two $t$–tests.
Lehmann versus \texttt{cendif}: Interpretation of results

- If $N_1 = N_2$, then there is no larger or smaller sample – and both methods work (especially \texttt{cendif}).
- If $N_1 < N_2$ (and $N_1$ is not too small), then the population variability of the smaller sample is best estimated using the sample variability of the smaller sample – favoring \texttt{cendif}.
- If $N_1 \ll N_2$ (and $N_1$ is very small), and we have prior reason to expect “similar” variability, then the population variability of the smaller sample is best estimated using the sample variability of the larger sample – favoring the Lehmann method.
- This seems to suggest a policy of regarding \texttt{cendif} as the default and the Lehmann formula as the “special case”, similar to the Moser–Stevens[2] policy regarding the two $t$–tests.
Possible further improvements to `cendif`

- The jackknife method used by `cendif` assumes $N_1 + N_2 - 1$ degrees of freedom, which may be over-generous if $N_1 \ll N_2$.
- It *might* be possible to devise an alternative degrees-of-freedom formula for the jackknife, like the Satterthwaite formula used in the unequal-variance $t$-test.
- The percentile bootstrap (Wilcox, 1998)\cite{wilcox1998} *might* possibly be an improvement on the `cendif` method.
- *However*, the 1000 subsamples typically used might make it computationally expensive to prove this in a study as large as this one!
Possible further improvements to cendif

- The jackknife method used by cendif assumes $N_1 + N_2 - 1$ degrees of freedom, which may be over-generous if $N_1 \ll N_2$.
- It might be possible to devise an alternative degrees-of-freedom formula for the jackknife, like the Satterthwaite formula used in the unequal-variance $t$-test.
- The percentile bootstrap (Wilcox, 1998) might possibly be an improvement on the cendif method.
- However, the 1000 subsamples typically used might make it computationally expensive to prove this in a study as large as this one!
Possible further improvements to cendif

- The jackknife method used by cendif assumes $N_1 + N_2 - 1$ degrees of freedom, which may be over-generous if $N_1 \ll N_2$.
- It *might* be possible to devise an alternative degrees-of-freedom formula for the jackknife, like the Satterthwaite formula used in the unequal-variance $t$-test.
- The percentile bootstrap (Wilcox, 1998)[5] *might* possibly be an improvement on the cendif method.
- *However*, the 1000 subsamples typically used might make it computationally expensive to prove this in a study as large as this one!
Possible further improvements to `cendif`

- The jackknife method used by `cendif` assumes $N_1 + N_2 - 1$ degrees of freedom, which may be over-generous if $N_1 \ll N_2$.
- It *might* be possible to devise an alternative degrees-of-freedom formula for the jackknife, like the Satterthwaite formula used in the unequal-variance $t$–test.
- The percentile bootstrap (Wilcox, 1998)[5] *might* possibly be an improvement on the `cendif` method.
- *However*, the 1000 subsamples typically used might make it computationally expensive to prove this in a study as large as this one!
Possible further improvements to \texttt{cendif}

- The jackknife method used by \texttt{cendif} assumes $N_1 + N_2 - 1$ degrees of freedom, which may be over-generous if $N_1 \ll N_2$.
- It \textit{might} be possible to devise an alternative degrees-of-freedom formula for the jackknife, like the Satterthwaite formula used in the unequal-variance $t$-test.
- The percentile bootstrap (Wilcox, 1998)[5] \textit{might} possibly be an improvement on the \texttt{cendif} method.
- \textit{However}, the 1000 subsamples typically used might make it computationally expensive to prove this in a study as large as this one!
Conclusions

- This simulation study compared the coverage probabilities of the Lehmann and `cendif` confidence intervals for median differences.
- Neither method failed “catastrophically”, in the manner of the t-test.
- *However*, both methods could be made to produce “95% confidence intervals” that were really 90% confidence intervals.
- Under *most* scenarios, it appears safe to use `cendif` as the default method.
- *However*, the Lehmann method *may* be better, if $N_1 \ll N_2$. 

Robust confidence intervals for Hodges-Lehmann median differences
This simulation study compared the coverage probabilities of the Lehmann and `cendif` confidence intervals for median differences.

Neither method failed “catastrophically”, in the manner of the $t$–test.

*However*, both methods could be made to produce “95% confidence intervals” that were really 90% confidence intervals.

Under *most* scenarios, it appears safe to use `cendif` as the default method.

*However*, the Lehmann method *may* be better, if $N_1 \ll N_2$. 
Conclusions

- This simulation study compared the coverage probabilities of the Lehmann and \texttt{cendif} confidence intervals for median differences.
- Neither method failed “catastrophically”, in the manner of the $t$–test.
- However, both methods could be made to produce “95% confidence intervals” that were really 90% confidence intervals.
- Under most scenarios, it appears safe to use \texttt{cendif} as the default method.
- However, the Lehmann method may be better, if $N_1 \ll N_2$. 
Conclusions

- This simulation study compared the coverage probabilities of the Lehmann and cendif confidence intervals for median differences.
- Neither method failed “catastrophically”, in the manner of the $t$–test.
- However, both methods could be made to produce “95% confidence intervals” that were really 90% confidence intervals.
- Under most scenarios, it appears safe to use cendif as the default method.
- However, the Lehmann method may be better, if $N_1 \ll N_2$. 
Conclusions

- This simulation study compared the coverage probabilities of the Lehmann and `cendif` confidence intervals for median differences.
- Neither method failed “catastrophically”, in the manner of the $t$–test.
- However, both methods could be made to produce “95% confidence intervals” that were really 90% confidence intervals.
- Under most scenarios, it appears safe to use `cendif` as the default method.
- However, the Lehmann method may be better, if $N_1 \ll N_2$. 

Robust confidence intervals for Hodges-Lehmann median differences
Conclusions

- This simulation study compared the coverage probabilities of the Lehmann and \texttt{cendif} confidence intervals for median differences.
- Neither method failed “catastrophically”, in the manner of the \textit{t}–test.
- \textit{However}, both methods could be made to produce “95% confidence intervals” that were really 90% confidence intervals.
- Under \textit{most} scenarios, it appears safe to use \texttt{cendif} as the default method.
- \textit{However}, the Lehmann method \textit{may} be better, if $N_1 \ll N_2$. 
References


This presentation can be downloaded from the conference website at http://ideas.repec.org/s/boc/usug07.html
Appendix

- This and the following frames are not part of the main presentation.
- However, they may be shown to the audience to illustrate responses to questions.
Median Gosset/confidence interval width ratios under equal variability

Graphs by Distributional family

Robust confidence intervals for Hodges-Lehmann median differences
Median Lehmann/cendif confidence interval width ratios under equal variability

Graphs by Distributional family

Robust confidence intervals for Hodges-Lehmann median differences
Normal coverage probabilities for the Gosset and cendif methods

<table>
<thead>
<tr>
<th>First/second population scale ratio</th>
<th>Gosset coverage</th>
<th>cendif coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>5, 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5, 10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5, 20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5, 40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10, 10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10, 20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10, 40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20, 20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20, 40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40, 40</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Graphs by First sample number and Second sample number

Robust confidence intervals for Hodges-Lehmann median differences
Cauchy coverage probabilities for the Gosset and cendif methods

Graphs by First sample number and Second sample number

- Gosset coverage
- cendif coverage

Coverage probability under Cauchy distribution

Robust confidence intervals for Hodges-Lehmann median differences
Normal coverage probabilities for the Satterthwaite and \texttt{cendif} methods

<table>
<thead>
<tr>
<th>First/second population scale ratio</th>
<th>Coverage probability under Normal distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>5, 5</td>
<td>![Graph for 5, 5]</td>
</tr>
<tr>
<td>5, 10</td>
<td>![Graph for 5, 10]</td>
</tr>
<tr>
<td>5, 20</td>
<td>![Graph for 5, 20]</td>
</tr>
<tr>
<td>5, 40</td>
<td>![Graph for 5, 40]</td>
</tr>
<tr>
<td>10, 10</td>
<td>![Graph for 10, 10]</td>
</tr>
<tr>
<td>10, 20</td>
<td>![Graph for 10, 20]</td>
</tr>
<tr>
<td>10, 40</td>
<td>![Graph for 10, 40]</td>
</tr>
<tr>
<td>20, 20</td>
<td>![Graph for 20, 20]</td>
</tr>
<tr>
<td>20, 40</td>
<td>![Graph for 20, 40]</td>
</tr>
<tr>
<td>40, 40</td>
<td>![Graph for 40, 40]</td>
</tr>
</tbody>
</table>

- **●** Satterthwaite coverage
- **◊** cendif coverage

Graphs by First sample number and Second sample number

Robust confidence intervals for Hodges-Lehmann median differences
Cauchy coverage probabilities for the Satterthwaite and cendif methods

<table>
<thead>
<tr>
<th>First/second population scale ratio</th>
<th>Coverage probability under Cauchy distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>5, 5</td>
<td><img src="image" alt="Graph" /></td>
</tr>
<tr>
<td>5, 10</td>
<td><img src="image" alt="Graph" /></td>
</tr>
<tr>
<td>5, 20</td>
<td><img src="image" alt="Graph" /></td>
</tr>
<tr>
<td>5, 40</td>
<td><img src="image" alt="Graph" /></td>
</tr>
<tr>
<td>10, 10</td>
<td><img src="image" alt="Graph" /></td>
</tr>
<tr>
<td>10, 20</td>
<td><img src="image" alt="Graph" /></td>
</tr>
<tr>
<td>10, 40</td>
<td><img src="image" alt="Graph" /></td>
</tr>
<tr>
<td>20, 20</td>
<td><img src="image" alt="Graph" /></td>
</tr>
<tr>
<td>20, 40</td>
<td><img src="image" alt="Graph" /></td>
</tr>
<tr>
<td>40, 40</td>
<td><img src="image" alt="Graph" /></td>
</tr>
</tbody>
</table>

Robust confidence intervals for Hodges-Lehmann median differences
Normal coverage probabilities for the Lehmann and *cendif* methods

<table>
<thead>
<tr>
<th>First/second population scale ratio</th>
<th>5, 5</th>
<th>5, 10</th>
<th>5, 20</th>
<th>5, 40</th>
<th>10, 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coverage probability under Normal distribution</td>
<td>0.25</td>
<td>0.3333</td>
<td>0.5</td>
<td>0.6667</td>
<td>1</td>
</tr>
<tr>
<td>First sample number</td>
<td>0.775</td>
<td>0.8</td>
<td>0.825</td>
<td>0.85</td>
<td>0.85</td>
</tr>
<tr>
<td>Second sample number</td>
<td>0.875</td>
<td>0.9</td>
<td>0.925</td>
<td>0.95</td>
<td>0.95</td>
</tr>
<tr>
<td>First/second population scale ratio</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Graphs by First sample number and Second sample number

Robust confidence intervals for Hodges-Lehmann median differences
Cauchy coverage probabilities for the Lehmann and *cendif* methods

Graphs by First sample number and Second sample number

Lehmann coverage  *cendif* coverage

Robust confidence intervals for Hodges-Lehmann median differences