

Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Copyright is retained by the first or sole author, who grants right of first publication to *Practical Assessment, Research & Evaluation*. Permission is granted to distribute this article for nonprofit, educational purposes if it is copied in its entirety and the journal is credited. PARE has the right to authorize third party reproduction of this article in print, electronic and database forms.

Volume 29 Number 2, January 2024

ISSN 1531-7714

Using Cliff's Delta as a Non-Parametric Effect Size Measure: An Accessible Web App and R Tutorial

Kane Meissel, *University of Auckland*
Esther S. Yao, *University of Auckland*

Effect sizes are important because they are an accessible way to indicate the practical importance of observed associations or differences. Standardized mean difference (SMD) effect sizes, such as Cohen's d , are widely used in education and the social sciences – in part because they are relatively easy to calculate. However, SMD effect sizes assume normally distributed data, whereas most data in these fields are ordinal and/or non-normal. In these situations, SMD effect sizes can be biased, and a non-parametric measure such as Cliff's delta (δ) is more appropriate. This paper provides a practical guide on how to calculate Cliff's δ . First, we present a conceptual overview and a worked example. Then we present two methods of calculating Cliff's δ : (1) a web-based Shiny application developed to accompany this paper (<https://clifffdelta.shinyapps.io/calculator>; suitable for all users), and (2) an R tutorial (suitable for R users). This is intended to provide researchers and practitioners with an appropriate and accessible effect size measure for non-normal data.

Keywords: Cliff's delta, effect size, non-parametric statistics, Shiny application, R

Introduction

Effect sizes quantify the magnitude of difference between two groups or time points, and indicate whether a relationship is of practical importance (Balkin & Lenz, 2021; Ferguson, 2009; Grissom & Kim, 2012). They are accessible to the wider educational community, and are often required or strongly suggested by large research organisations and publication outlets, especially when statistical significance is misleading, for example when sample sizes are small or very large (American Educational Research Association, 2006; American Psychological Association, 2020; Peng et al., 2013). Effect size metrics fall into three main families: odds ratios, correlations, and standardized mean difference measures (Hedges, 2008). In educational research and evaluation, effect sizes are most commonly based on standardised mean difference (SMD), such as Cohen's

(1988) d , arguably due to ease of calculation (Farmus et al., 2022; Ferguson, 2009; Fritz et al., 2012; Peng et al., 2013). SMD effect sizes are calculated by dividing the difference in means between two groups by the standard deviation (with some small variations in terms of which standard deviation and/or how to combine these). However, unbiased SMD effect sizes require several assumptions to be met, including the data to be measured on a metric scale, be normally distributed, and have homogeneity of variance (Cohen, 1988; Grissom & Kim, 2012; Hedges & Olkin, 1985). These assumptions are rarely satisfied in data in education and the social sciences (Grissom & Kim, 2012; Micceri, 1989; Romano et al., 2006). Therefore, means and standard deviations are frequently a poor representation of the distributional properties of the data, and can lead to inaccurate and unstable effect size results (Algina et al., 2005; Hess & Kromrey, 2004; Marfo & Okyere, 2019).

In cases where the aforementioned assumptions are *not* met, non-parametric effect sizes, such as Cliff's (1993, 1996) delta (δ), produce substantially more robust and reliable results (Grissom & Kim, 2012; Hess & Kromrey, 2004; Romano et al., 2006). This is because non-parametric effect sizes are suitable for ordinal data, and do not make assumptions about underlying distributional properties. There are no drawbacks to using non-parametric effect sizes even when normality assumptions are met, other than traditionally having been more complex to calculate (Cliff, 1993, 1996). Non-parametric effect sizes are thus rarely used, presumably because there is less information about these effect sizes and how to calculate them, coupled with the inability of most major statistical packages to compute it for users (Grissom & Kim, 2012; Ledesma et al., 2009; Romano et al., 2006). The major motivation for this paper is to make Cliff's δ more accessible, via a practical guide for researchers and practitioners. As Cliff's δ remains relatively rarely used, an introduction to the metric as a viable non-parametric effect size measure is provided first, followed by presentation of a novel web-based Shiny application developed by the authors (<https://cliffdelta.shinyapps.io/calculator>; suitable for all users) and an R tutorial (suitable for users of the R statistical software).

Cliff's Delta as a Non-Parametric Effect Size

Cliff's (1993, 1996) delta (δ) is a robust and intuitive non-parametric alternative to SMD effect sizes such as Cohen's d . It is especially useful when data are non-normal, or are ordinal and hence have reduced variance (e.g., Likert scale responses from surveys). Cliff's δ was originally conceptualised as a dominance statistic, and is obtained by calculating the non-overlapping area of two distributions (e.g., School 1 vs. School 2; Time 1 vs. Time 2) at the individual case level. The statistic ranges from -1 to $+1$, with the extremes indicating no overlap between the two distributions, and 0 indicating complete overlap. One major advantage of δ is that it measures the magnitude of an effect across the entire distributions of both groups, rather than simply the effect for those in the centre of the distribution, which is a common criticism of SMD effect sizes (Grissom & Kim, 2012).

While conceptually simple, measuring the proportion of distributional non-overlap for Cliff's δ is a relatively involved task requiring matrix statistics. In short, a matrix is constructed to determine the proportion of instances where values from one group are larger than values from another (i.e., the ratio of "dominance" of one group over the other). The calculation of Cliff's δ is formalised as:

$$\text{Cliff's delta } (\delta) = \frac{\#(x_{i1} > x_{j2}) - \#(x_{i1} < x_{j2})}{n_1 n_2} \quad (1)$$

In this equation, the number of instances (#) where a case in Group 1 (x_{i1}) is larger than Group 2 (x_{j2}) is subtracted by the number of instances (#) where a case in Group 1 (x_{i1}) is smaller than Group 2 (x_{j2}). The difference is then divided by the product of the sample sizes of Group 1 (n_1) and Group 2 (n_2). All possible comparisons are made, meaning that the entire distribution (i.e., all individuals from both groups) is represented by the statistic. The choice of which group is subtracted from the other affects whether the statistic is positive or negative, but will not affect the absolute value of the result.

For example, consider a small fictitious dataset with two dependent groups (Time 1 and Time 2). The scores for Time 1 ($n = 6$) are 1, 2, 2, 3, 3, and 4; and the scores for Time 2 ($n = 5$) are 2, 3, 3, 4, and 4. Table 1 shows the dominance matrix for the data, where each score from Time 1 is compared to each score from Time 2. The value 1 is assigned when the score in Time 1 is larger than Time 2; 0 is assigned when the scores are equal; and -1 is assigned when the score in Time 1 is smaller than Time 2. The sum of these values (-12) divided by the product of the group sizes ($6 \times 5 = 30$) results in a Cliff's δ effect size of -0.40 . The result is negative, meaning the distribution of Time 1 is smaller than the distribution of Time 2.

In our view, this directionality is somewhat unintuitive because it treats Time 2 as the reference group rather than Time 1. To resolve this, we can simply reverse the order of the groups so that Time 1 is treated as the reference group. This reverses the signs in the dominance matrix, resulting in a sum of 12 and a Cliff's δ of 0.40 . The result is positive, meaning the distribution of Time 2 is larger than the distribution of Time 1. We believe this directionality is more intuitive to most researchers and practitioners, so we will treat Time 1 – or Group 1 for independent groups – as the reference group in the tutorial section of this paper.

Interpreting Cliff's Delta

Cliff's δ can be interpreted as the degree of distributional non-overlap between two distributions (Cliff, 1993, 1996). For instance, using the previous example, $\delta = 0.40$ indicates a 40% non-overlap (or 60% overlap) between Time 1 and Time 2. Cliff's δ can also be interpreted by converting it to the more common Cohen's d metric, since Cohen's d can also be expressed as a measure of distributional non-overlap (provided the data are normally distributed; Cohen, 1988). For example, a 33% non-overlap is equivalent to a Cohen's d of 0.50 (for a conversion table, see Table 2.2.1 in Cohen, 1988). Therefore, using the degree of non-overlap (33%) as a "bridge", a Cliff's δ of 0.33 is equivalent to a Cohen's d of 0.50. Since Cliff's δ does not make distributional assumptions, this bridge allows conversion of δ into an unbiased estimate of Cohen's d for non-normal data (Romano et al., 2006). The discrepancy between Cohen's d derived from Cliff's δ (via the bridge), and Cohen's d calculated from raw data

(mean difference divided by standard deviation), demonstrates the level of bias attributable to using raw Cohen's d for non-normal data.

It is also common to interpret effect sizes as "negligible", "small", "medium", or "large". Table 2 shows the benchmarks for interpreting Cohen's d as suggested by Cohen (1988), as well as the equivalent interpretation for Cliff's δ (calculated via the bridge). Note that these are only conventional rules of thumb. It is best to interpret effect sizes in light of previous research in the relevant field of investigation wherever possible (Balkin & Lenz, 2021; Ferguson, 2009; Grissom & Kim, 2012).

Missing Data and Dependency Issues

Cliff's δ compares each case in Group 1 to each case in Group 2, so missing data are not an issue for independent groups (e.g., control vs. intervention; male vs. female)—the corresponding comparisons are simply not made. However, care is needed when calculating Cliff's δ from dependent, or paired, data

Table 1. Example Dominance Matrix

		Time 2 scores				
		2	3	3	4	4
Time 1 scores	2	-1	-1	-1	-1	-1
1	0	-1	-1	-1	-1	-1
2	0	-1	-1	-1	-1	-1
3	1	0	0	-1	-1	-1
3	1	0	0	-1	-1	-1
4	1	1	1	0	0	0

Table 2. Effect Size Interpretation

Interpretation	Cohen's d	Cliff's delta (δ)
Negligible	<0.20	<0.15
Small	0.20	0.15
Medium	0.50	0.33
Large	0.80	0.47

Note. Cliff's delta (δ) was converted from Cohen's d using the degree of non-overlap between two distributions as a bridge.

(e.g., pre-intervention vs. post-intervention; beginning-of-year vs. end-of-year). When working with dependent data, it is helpful to consider what type of change is of interest (e.g., overall change, within-person change, or between-person change), as well as which participants are of interest (e.g., should participants without Time 2 data be included or excluded?).

Previous research notes that researchers frequently fail to indicate whether dependency has been accounted for, so there is a considerable threat to the validity of meta-analyses that combine effect sizes derived from different research designs since the various approaches have the potential to produce large differences in the apparent effect size magnitude (Ahn et al., 2013; Morris & DeShon, 2002). In general, those based on within-person change (i.e., matched comparisons) produce larger effect sizes (Dunlap et al., 1996). In alignment with Dunlap et al., (1996), our position is that while there is increased reliability in the statistical estimates associated with matched comparisons, there is generally no justification for adjusting the magnitude of the effect, unless specifically appropriate for the research question (e.g., special education with a small number of individuals; Morris & DeShon, 2002). For those who need to do so, calculation of within-person Cliff's δ is relatively straight-forward, and a description is included in the footnote below.¹

Researchers and practitioners analysing dependent data in education and the social sciences interested in the effect of time (e.g., the effect of an intervention on achievement) need to calculate Cliff's δ for the *overall distribution*, or in other words, the overall direction and extent to which participants have shifted over time. To examine how an *overall distribution* has changed over time, Cliff's δ is calculated in the same way as independent data (i.e., all comparisons in the dominance matrix are included). When using this method for dependent data, users still need to consider whether to include or exclude participants who had data at Time 1 but not at Time 2 (e.g., due to attrition,

unavailability, etc.). Cliff's δ can be calculated with missing data, so the decision depends on the substantive question(s) of interest. When interpreting results where the sample is only partially matched, it is important to acknowledge that the effect size reflects a more descriptive evaluation of any difference in classroom performance at the two time points rather than the change over time for individual students – this is exacerbated by the extent of missingness in the sample. In contrast, if the substantive question is about the effectiveness of an intervention or the *change* for individual students, consideration of the overall distributional change of the matched sample is usually more appropriate, though the retention rate and any attrition bias needs to be considered (Rosenbaum & Rubin, 1985).

Example Dataset and Results

To illustrate how to calculate Cliff's δ , we will utilise an example dataset of overall teacher judgments (OTJs) of student achievement (see Figure 1 in the next section for the first ten rows of data; full dataset available in Supplementary Materials). The dataset contains two columns of data corresponding to School 1 and School 2, but the principles and steps described apply to any two independent or dependent groups. The OTJs are rated on a four-point ordinal scale in relation to curriculum expectations for the student's year level (1 = *well below standard*, 2 = *below standard*, 3 = *at standard*, and 4 = *above standard*). The ordinal nature of the data means Cliff's δ is a more suitable effect size measure than Cohen's d .

Table 3 shows the effect size results and related statistics for the example dataset. The first two rows show the number of cases from each school (note School 1 had more cases than School 2). The next three rows show the probability that a case in School 1 scored higher than, equal to, and lower than School 2. The probability that a case in School 1 scored higher than School 2 (0.27) is smaller than the probability that a case in School 1 scored lower than School 2 (0.40).

¹ To calculate within-person change (δ_w), simply calculate the proportion of cases where *matched* Time 2 scores are higher than Time 1 scores, minus the proportion of cases where *matched* Time 1 scores are higher than Time 2 scores (Cliff, 1996). For example, in a case with five participants where three have higher scores at Time 2 (3/5 = .6), one has a lower score (1/5 = .2) and one has the same score (not included in the calculation), the within-person change is $.6 - .2 = .4$.

In addition, 33% of the scores were equal between the two schools. Overall, this indicates that OTJs in School 2 are generally higher than School 1, but it is hard to evaluate the magnitude of this difference. The following row shows the calculated Cliff's δ effect size difference between the two schools (0.12) is approaching small in magnitude. The final two rows show the discrepancy between using a non-parametric effect size measure (Cohen's d converted from Cliff's δ using the proportion of non-overlap as a bridge = 0.16) and a SMD effect size measure (Cohen's d calculated from raw data, with Hedges' bias correction² = 0.31) for a dataset that violates parametric assumptions. In this case, the SMD measure overestimates the effect size by $d = 0.15$. The exact level of discrepancy will differ depending on the dataset analysed, indicating the importance of using a non-parametric effect size such as Cliff's δ if parametric assumptions are not met.

Calculating Cliff's Delta

The following tutorial demonstrates two methods to calculate Cliff's (1993, 1996) δ and the accompanying statistics shown in Table 3. The first

method is via a freely accessible web-based Shiny application developed to accompany this paper (<https://cliffdelta.shinyapps.io/calculator>). To obtain results, users simply need to upload an appropriate dataset to the app. This method is suitable for all users. The second method is via the R open-source statistical software (R Core Team, 2022). This involves entering code into the command line, and is primarily targeted at users with basic working knowledge of R. The tutorial will first outline how to format a dataset for analysis. The same data format is required for both the Shiny app and R.

Data Preparation

To calculate Cliff's δ in the Shiny app or R, the dataset needs to be structured as two columns in a spreadsheet (e.g., Microsoft Excel), with Column 1 containing numeric data for Group 1 (or Time 1), and Column 2 containing numeric data for Group 2 (or Time 2). The first row of the spreadsheet should indicate the name of the group corresponding to each column. The names should begin with a letter but can contain numbers (e.g., School1, School2); no spaces or special characters are allowed. The actual data begins from the second row. Cells with missing data should be left blank. If you are working with dependent data

Table 3. Effect Size Results for Example Dataset

Statistic	Value	Lower 95% CI	Upper 95% CI
Number of cases in School 1	30.00	-	-
Number of cases in School 2	25.00	-	-
Probability that School 1 > School 2	0.27	-	-
Probability that School 1 = School 2	0.33	-	-
Probability that School 1 < School 2	0.40	-	-
Cliff's δ (calculated from raw data)	0.12	-0.17	0.39
Cohen's d (converted from Cliff's δ)	0.16	-0.20	0.62
Cohen's d ^a (calculated from raw data)	0.31	-0.23	0.85

Note. CI = confidence interval (see e.g., Hess & Kromrey, 2004, for details on how the CI is calculated).

^aWith Hedges' correction; pooled standard deviation used.

² Also known as Hedges' g . Hedges' g corrects the positive bias in Cohen's d . It is calculated by multiplying Cohen's d with a correction factor of $J = 1 - \frac{3}{4(n_1+n_2)-9}$ (Hedges & Olkin, 1985).

and wish to exclude participants with missing data at either time point (see *Missing Data and Dependency Issues*), the available data for these participants need to be deleted (this typically involves removing the entire row of data corresponding to the relevant participants). The prepared dataset needs to be saved as a comma-separated values (CSV) file.³ Figure 1 shows the first ten rows of the example dataset (see *Example Dataset and Results*) as prepared for analysis (the complete prepared dataset is downloadable from Supplementary Materials). If using the Shiny web app, the dataset should not exceed 5000 rows.

Using the Shiny Web App

A novel Shiny app was developed for this paper as a freely accessible web-based tool to calculate Cliff's δ and associated statistics.⁴ A screenshot of the app's input panel is shown in Figure 2. To use the app, simply navigate to <https://cliffdelta.shinyapps.io/calculator> (an Internet connection is required) and upload the prepared CSV

file (see *Data Preparation*) by selecting "Browse..." in the input panel. Alternatively, users can tick the box next to "Use example data" to calculate the results for the example dataset previously described in this paper. The column names provided in the CSV file are used by default in the outputted results. If needed, users can rename the two columns via the input panel (letters, numbers, spaces, and special characters are all allowed).

Once a CSV file is uploaded (or the example data option is ticked), the results will be shown in a table under the *Results* heading. The statistics calculated are the same as those shown in Table 3 (for a written explanation of these statistics, see the *Example Dataset and Results* section in this paper). We strongly recommend checking that the number of cases displayed in the outputted table is consistent with the uploaded dataset. Under the table, there is a button that can be used to download the results as a CSV file. A written interpretation of the results is available under

Figure 1. First Ten Rows of Example Dataset

	A	B
1	School1	School2
2	3	4
3	1	2
4	4	4
5		4
6	3	3
7	3	
8	2	
9	2	2
10	4	3

³ To save a Microsoft Excel file in CSV format, open the relevant Excel file, then click "File" followed by "Save As". In the popup window, click the dropdown list next to "Save as type", then select "CSV UTF-8 (Comma-delimited) (.csv)". For more detailed instructions, see <https://support.microsoft.com/en-us/office/save-a-workbook-to-text-format-txt-or-csv-3e9a9d6c-70da-4255-aa28-fcacf1f081e6>.

⁴ The Shiny app was built using the *Shiny* package (version 1.7.1; Chang et al., 2020) in R (version 4.1.3; R Core Team, 2022), and depends on the *effsize* package (version 0.8.1; Torchiano, 2020) to compute raw Cliff's δ and Cohen's d . The app's underlying code is identical to that shown in the R tutorial section (all modifications are purely aesthetic and to broaden accessibility).

Figure 2. Screenshot of Input Panel in the Shiny App

The screenshot shows the input panel of a Shiny application. At the top, there is a button labeled "Upload CSV file:" with a "Browse..." button and a text field showing "No file selected". Below this, a section titled "Or tick to use example data:" contains a checkbox labeled "Use example data" which is unchecked. Underneath, there are two sections for renaming columns: "Rename Column 1:" with a text field containing "(optional)" and "Rename Column 2:" with another text field containing "(optional)".

the *Interpretation* heading. An error message in the results and/or interpretation section usually suggests that the uploaded dataset is incorrectly formatted. Please follow the instructions in this paper's *Data Preparation* section to resolve the issue.

Using R Statistical Software

Users with working knowledge of the R statistical software (R Core Team, 2022) can also calculate Cliff's δ using the *effsize* package (Torchiano, 2020). Those who do not wish to calculate Cliff's δ using R can skip this section. *Effsize* contains functions to calculate both Cliff's δ and SMD effect sizes (e.g., Cohen's d). The tutorial below provides a step-by-step guide to using R to calculate Cliff's δ and the associated statistics shown in Table 3. It assumes that users have R (and RStudio if desired) installed on their computer.⁵ A downloadable version of the R code used below is available in Supplementary Materials.

Step 1: Install and Load the *effsize* Package. To begin, the *effsize* package (Torchiano, 2020) needs to be installed (if it has not already been installed previously). To do this, call:

```
install.packages("effsize")
```

Then, use the `library()` function to load the *effsize* package into the current R session (note this needs to be loaded every time a new session is started):

```
library(effsize)
```

An overview of the *effsize* package is available via the command `help(effsize)`. This tutorial is based on *effsize* version 0.8.1, the current version at the time of writing.

Step 2: Load the Prepared Dataset. Next, the dataset prepared according to the instructions in *Data Preparation* needs to be loaded into the R environment. This should be a CSV file with two columns of data, and should contain a header row. Load the dataset as

⁵ R can be downloaded for free from <https://cran.r-project.org/>. We recommend using R within RStudio, an integrated development environment that makes working with R easier. RStudio can be downloaded for free from <https://posit.co/download/rstudio-desktop/>.

follows (change the file path to where your CSV file is located):

```
df <-  

read.csv("C:/Users/UserName/Documents/Data/example-data.csv",  

        fileEncoding = "UTF-8-BOM")
```

You can view the first six rows of the loaded dataset by calling `head(df)`:

```
> head(df)  

  School1 School2  

1       3       4  

2       1       2  

3       4       4  

4      NA       4  

5       3       3  

6       3      NA
```

Step 3: Calculate Cliff's Delta. To calculate the Cliff's δ effect size difference between the dataset's two columns, use the `cliff.delta()` function provided by the `effsize` package, and replace the column names shown in parentheses below with your column names:

```
cliff.delta(df$Column2, df$Column1)
```

The first column specified is treated as the comparison group, and the second column as the reference group. We have specified Column 2 first because, in our view, treating Column 2 as the comparison group provides more intuitive results. Note that, as explained earlier, the order of the columns affects whether the result is positive or negative, but does not affect the absolute Cliff's δ result.

We can obtain the Cliff's δ estimate and 95% CI for the example dataset as follows:

```
> cliff.delta(df$School2,  

df$School1)  

Cliff's Delta  

delta estimate: 0.1213333  

(negligible)  

95 percent confidence interval:  

  lower      upper
```

-0.1694430 0.3926771

Here, the result is positive. Because School 2 was specified as the comparison group, this indicates the distribution of School 2 is larger than that of School 1 (if School 1 was specified as the comparison group, the results would be $\delta = -0.12$, 95% CI $[-0.39, 0.17]$). The output also provides a written interpretation (i.e., negligible). The interpretation is based on the same benchmarks as those shown in Table 2.

Step 4: Calculate Additional Statistics (Optional).

In some cases, users may like to obtain the additional statistics shown in Table 3. We provide instructions for this below.

Number of Cases. To calculate the number of valid cases in a column, call:

```
length(na.omit(df$ColumnName))
```

This counts the number of cases with a non-missing value in the specified column. For example, the respective number of non-missing cases for School 1 and School 2 in the example dataset is:

```
> length(na.omit(df$School1))  

[1] 30  

> length(na.omit(df$School2))  

[1] 25
```

Probability of Superiority. To calculate the probabilities that a case in Column 1 is larger than, equal to, and smaller than Column 2, we need to compute the dominance matrix. This is done via the previously introduced `cliff.delta()` function. To aid interpretability, we again specify Column 2 as the first argument (treated as the comparison group), and Column 1 as the second argument (treated as the reference group). However, this time, we also need to specify `return.dm = TRUE` to obtain the dominance matrix, and save the result as an object:

```
result <- cliff.delta(df$Column2,  

df$Column1, return.dm = TRUE)
```

The dominance matrix can be accessed via `result$dm`. We will use this below.

To calculate the probability that a value in Column 1 is *larger* than Column 2 (given that Column 2 was specified as the first argument), call:

```
sum(result$dm <
0)/length(result$dm)
```

This sums the number of times that a value in Column 1 is larger than Column 2, and divides it by the number of comparisons made (i.e., size of the dominance matrix). Using the same logic, the probability that a value in Column 1 *equals* Column 2 is calculated via:

```
sum(result$dm ==
0)/length(result$dm)
```

The probability that a value in Column 1 is *smaller* than Column 2 calculated via:

```
sum(result$dm >
0)/length(result$dm)
```

Cohen's *d* Converted from Cliff's Delta. To convert Cliff's δ (and its 95% CI) to Cohen's d metric using the proportion of non-overlap as a bridge, we need to use the formula $2 * \text{qnorm}(-1/(\delta-2))$, where `qnorm()` is a function that returns the inverse of the cumulative density function, and δ is the Cliff's δ value we are wanting to convert from (Cohen, 1988; Rogmann, 2013). To do so in a replicable way, first save the Cliff's δ results as an object:

```
result <- cliff.delta(df$Column2,
df$Column1)
```

Then, convert the saved Cliff's δ value (accessed via `result$estimate`), as well as its lower 95% CI (accessed via `result$conf.int[[1]]`) and upper 95% CI (accessed via `result$conf.int[[2]]`), to Cohen's d using the stated formula.

Here is the corresponding code and Cohen's d results for the example dataset:

```
> result <- cliff.delta(df$School2,
df$School1)
> 2*qnorm(-1/(result$estimate-2)) # Cohen's d estimate
[1] 0.1620673
> 2*qnorm(-1/(result$conf.int[[1]] -
```

```
2)) # Lower 95% CI for Cohen's d estimate
[1] -0.1960924
> 2*qnorm(-1/(result$conf.int[[2]] - 2)) # Upper 95% CI for Cohen's d estimate
[1] 0.622278
```

The text after the `#` character specifies what is computed in each line of code (R treats text after the `#` character as a comment and does not execute it). Hence, for our example, Cliff's δ results converted to Cohen's d metric is $d = 0.16$, 95% CI [-0.20, 0.62].

Cohen's *d* Calculated from Raw Data. To examine the discrepancy between using a non-parametric effect size measure (Cohen's d converted from Cliff's δ) and a SMD effect size measure (Cohen's d calculated from raw data), users can calculate Cohen's d from raw data using the `cohen.d()` function in the `effsize` package. The function is called in a similar way to the `cliff.delta()` function:

```
cohen.d(df$Column2, df$Column1,
pooled = TRUE, hedges = TRUE,
na.rm = TRUE)
```

Similarly, the output shows the Cohen's d estimate, the 95% CI, and a written interpretation (e.g., small).

As with previously, we have specified Column 2 as the first argument in the command, so that it is treated as the comparison group for interpretability reasons. Again, the order of the columns only affects whether the results are positive or negative; it does not affect the absolute Cohen's d value. We have also specified `pooled = TRUE` so that the calculation uses the pooled standard deviation of the two columns,⁶ `hedges = TRUE` to request Hedges' bias correction (see Footnote 1), and `na.rm = TRUE` to remove missing values from the calculation. Users can change these settings to `FALSE` if needed. Note that if `pooled = FALSE`, the standard deviation of the column specified in the second argument (e.g., `df$Column1`) is used. More information about these options can be accessed by calling `?cohen.d` in R.

⁶ The formula used for pooled standard deviation is $s = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$.

Limitations

This paper presents Cliff's δ as a non-parametric effect size measure that is more accurate and appropriate than SMD effect size measures (e.g., Cohen's d) when data are ordinal, continuous but non-normal, and/or have heterogeneity of variance. As discussed, it can be used in almost any case where the difference in two distributions is being assessed. Nevertheless, some limitations of this measure should be noted. Specifically, the measure is currently restricted to bivariate comparisons, meaning that researchers cannot easily include other covariates that might be influencing differences in the distributional overlap (e.g., socioeconomic status). In addition, due to the computational requirements inherent in calculating the dominance matrix, particularly large distributions (e.g., sample sizes above ~ 5000) remain slow to compute despite considerable improvements over the past decade. Nevertheless, in cases where parametric effect sizes are used despite violation of normality assumptions, Cliff's δ provides a more accurate alternative.

Conclusions

This paper provides a conceptual explanation and practical example, followed by a practical guide on how to calculate Cliff's δ using a novel web-based Shiny app (suitable for all users) and using R (suitable for R users). The paper is intended to provide researchers and practitioners working with non-normal data with an accessible non-parametric alternative to SMD effect sizes.

References

- Ahn, S., Ames, A. J., & Myers, N. D. (2012). A review of meta-analyses in education: Methodological strengths and weaknesses. *Review of Educational Research*, 82, 436-476. <https://doi.org/10.3102/0034654312458162>
- Algina, J., Keselman, H. J., & Penfield, R. D. (2005). An alternative to Cohen's standardized mean difference effect size: A robust parameter and confidence interval in the two independent groups case. *Psychological Methods*, 10(3), 317-328. <https://doi.org/10.1037/1082-989X.10.3.317>
- American Educational Research Association. (2006). Standards for reporting on empirical social science research in AERA publications. *Educational Researcher*, 35(6), 33-40. <https://doi.org/10.3102/0013189X035006033>
- American Psychological Association. (2020). *Publication Manual of the American Psychological Association: The official guide to APA style* (7th ed.). American Psychological Association.
- Balkin, R. S., & Lenz, A. S. (2021). Contemporary issues in reporting statistical, practical, and clinical significance in counseling research. *Journal of Counseling & Development*, 99(2), 227-237. <https://doi.org/10.1002/jcad.12370>
- Chang, W., Cheng, J., Allaire, J., Xie, Y., & McPherson, J. (2020). *shiny: Web application framework for R*. <https://cran.r-project.org/package=shiny>
- Cliff, N. (1993). Dominance statistics: Ordinal analyses to answer ordinal questions. *Psychological Bulletin*, 114(3), 494-509. <https://doi.org/10.1037/0033-2909.114.3.494>
- Cliff, N. (1996). *Ordinal methods for behavioral data analysis*. Lawrence Erlbaum Associates.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods*, 1(2), 170-177.
- Farmus, L., Beribisky, N., Martinez Gutierrez, N., Alter, U., Panzarella, E., & Cribbie, R. A. (2022). Effect size reporting and interpretation in social personality research. *Current Psychology*. <https://doi.org/10.1007/s12144-021-02621-7>
- Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice*, 40(5), 532-538. <https://doi.org/10.1037/a0015808>
- Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, 141(1), 2-18. <https://doi.org/10.1037/a0024338>

- Grissom, R. J., & Kim, J. J. (2012). *Effect sizes for research: Univariate and multivariate applications* (2nd ed.). Routledge.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic Press.
- Hedges, L. V. (2008). What are effect sizes and why do we need them?. *Child development perspectives*, 2(3), 167-171.
- Hess, M. R., & Kromrey, J. D. (2004, April 12–16). *Robust confidence intervals for effect sizes: A comparative study of Cohen's d and Cliff's delta under non-normality and heterogeneous variances* [Paper presentation]. Annual Meeting of the American Educational Research Association, San Diego, CA, United States
- Leedesma, R. D., Macbeth, G., & Cortada de Kohan, N. (2009). Computing effect size measures with ViSta - The Visual Statistics System. *Tutorials in Quantitative Methods for Psychology*, 5(1), 25–34. <https://doi.org/10.20982/tqmp.05.1.p025>
- Marfo, P., & Okyere, G. A. (2019). The accuracy of effect-size estimates under normals and contaminated normals in meta-analysis. *Helijon*, 5, 1–9. <https://doi.org/10.1016/j.heliyon.2019.e01838>
- Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, 105(1), 156–166. <https://doi.org/10.1037/0033-2909.105.1.156>
- Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods*, 7, 105-25. <https://doi.org/10.1037/1082-989x.7.1.105>
- Peng, C.-Y. J., Chen, L.-T., Chiang, H.-M., & Chiang, Y.-C. (2013). The impact of APA and AERA guidelines on effect size reporting. *Educational Psychology Review*, 25(2), 157–209. <https://doi.org/10.1007/s10648-013-9218-2>
- R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.r-project.org>
- Rogmann, J. J. (2013). *orddom: Ordinal dominance statistics*. <https://cran.r-project.org/package=orddom>
- Romano, J., Kromrey, J. D., Coraggio, J., Skowronek, J., & Levine, L. (2006, October 14–17). *Exploring methods for evaluating group differences on the NSSE and other surveys: Are the t-test and Cohen's d indices the most appropriate choices?* [Paper presentation]. Annual Meeting of the Southern Association for Institutional Research, Arlington, VA, United States.
- Rosenbaum, P. R., & Rubin, D. B. (1985). The bias due to incomplete matching. *Biometrics*, 103-116. <https://doi.org/10.2307/2530647>
- Torchiano, M. (2020). *effsize: Efficient effect size computation*. <https://cran.r-project.org/package=effsize>

Citation:

Meissel, K., & Yao, E. S. (2024). Using Cliff's Delta as a non-parametric effect size measure: An accessible web app and R tutorial. *Practical Assessment, Research, & Evaluation*, 29(2). Available online: <https://doi.org/10.7275/pare.1977>

Corresponding Author:

Kane Meissel
University of Auckland

Email: k.meissel [at] auckland.ac.nz

Supplementary Materials

1. Example dataset (.csv format)
2. Annotated R code (.R format)