Exploring Equivalence Testing with the TOSTER R package

Aaron R. Caldwell

Abstract

This is an article detailing the “avocado TOST” update to the TOSTER R package.

*Keywords:* statistics, bootstrapping, minimal effects test

Exploring Equivalence Testing with the TOSTER R package

# 1 Introduction

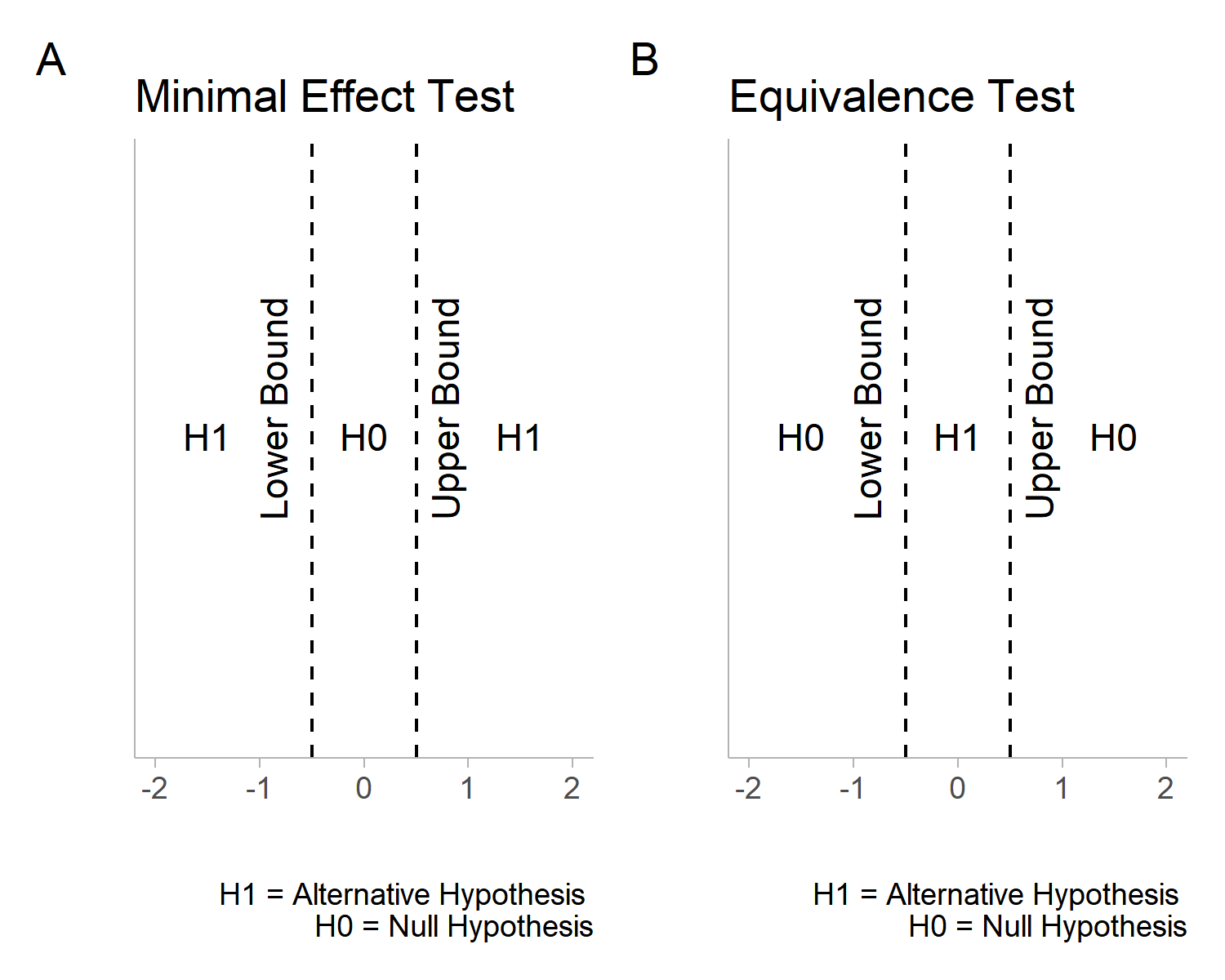
Importance of equivalence testing Seems to be mostly unrecognized by general researchers Many of the complaints about significance testing can be addressed through equivalence testing Teaching the average researcher about equivalence testing, and providing easy-to-use tools to implement such testing, could improve awareness TOSTER (Avocado TOST edition >v0.4.0) is a good tool to introduce the idea of equivalence testing to the average researcher who is already familiar with equivalence testing.

# 2 Basics of Equivalence Testing

## 2.1 The TOSTER R Package

In an effort to make TOSTER more informative and easier to use, a new function t\_TOST has been created. This function operates very similarly to base R’s t.test function, but performs 3 t-tests (one two-tailed and two one-tailed tests). In addition, this function has a generic method where two vectors can be supplied or a formula can be given (e.g.,y ~ group). This function also makes it easier to switch between types of t-tests. All three types (two sample, one sample, and paired samples) can be performed/calculated from the same function. Moreover, the summary information and visualizations have been upgraded. This should make the decisions derived from the function more informative and user-friendly.

Also, t\_TOST is not limited to equivalence tests. Minimal effects testing (MET) is now possible. MET is useful for situations where the hypothesis is about a minimal effect and the null hypothesis *is* equivalence.



In the general introduction to this package we detailed how to look at *old* results and how to apply TOST to interpreting those results. However, in many cases, users may have new data that needs to be analyzed. Therefore, t\_TOST can be applied to new data. This vignette will use the bugs data from the jmv R package and the sleep data.

data('sleep')  
library(jmv)  
data('bugs')

# 3 Independent Groups

For this example, we will use the sleep data. In this data there is a group variable and an outcome extra.

head(sleep)

## extra group ID  
## 1 0.7 1 1  
## 2 -1.6 1 2  
## 3 -0.2 1 3  
## 4 -1.2 1 4  
## 5 -0.1 1 5  
## 6 3.4 1 6

We will assume the data are independent, and that we have equivalence bounds of +/- 0.5. All we need to do is provide the formula, data, and eqbound arguments for the function to run appropriately. In addition, we can set the var.equal argument (to assume equal variance), and the paired argument (sets if the data is paired or not). Both are logical indicators that can be set to TRUE or FALSE. The alpha is automatically set to 0.05 but this can also be adjusted by the user. The Hedges correction is also automatically calculated, but this can be overridden with the bias\_correction argument. The hypothesis is automatically set to “EQU” for equivalence but if a minimal effect is of interest then “MET” can be supplied

res1 = t\_TOST(formula = extra ~ group,  
 data = sleep,  
 low\_eqbound = -.5,  
 high\_eqbound = .5)  
  
res1a = t\_TOST(x = subset(sleep,group==1)$extra,  
 y = subset(sleep,group==2)$extra,  
 low\_eqbound = -.5,  
 high\_eqbound = .5)

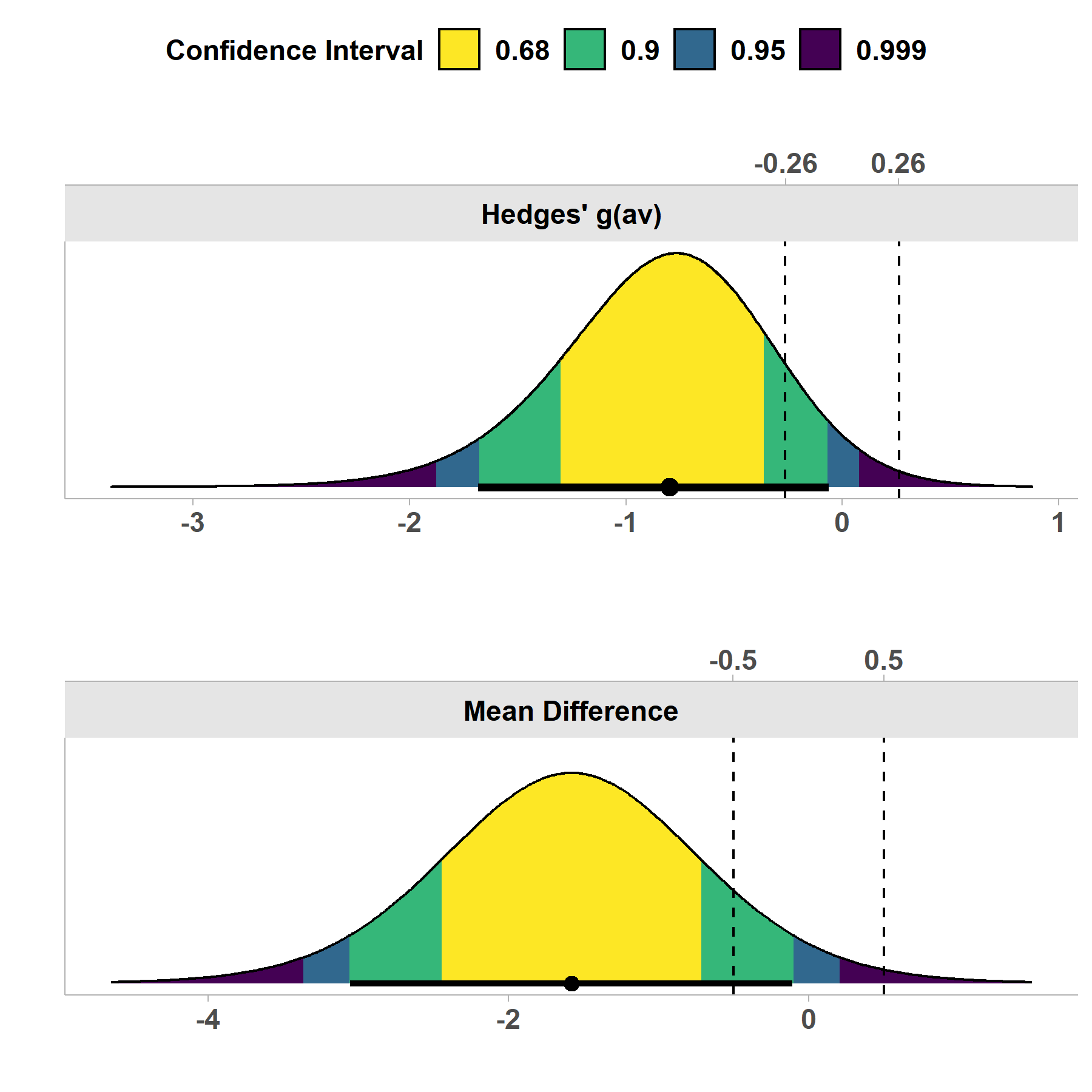
Once the function has run, we can print the results with the print command. This provides a verbose summary of the results.

print(res1)

##   
## Welch Two Sample t-test  
## Hypothesis Tested: Equivalence  
## Equivalence Bounds (raw):-0.500 & 0.500  
## Alpha Level:0.05  
## The equivalence test was non-significant, t(17.78) = -1.272, p = 8.9e-01  
## The null hypothesis test was non-significant, t(17.78) = -1.861, p = 7.94e-02  
## NHST: don't reject null significance hypothesis that the effect is equal to zero   
## TOST: don't reject null equivalence hypothesis  
##   
## TOST Results   
## t SE df p.value  
## t-test -1.860813 0.849091 17.77647 0.07939414  
## TOST Lower -1.271948 0.849091 17.77647 0.89010996  
## TOST Upper -2.449678 0.849091 17.77647 0.01245133  
##   
## Effect Sizes   
## estimate SE lower.ci upper.ci conf.level  
## Raw -1.5800000 0.849091 -3.053381 -0.10661850 0.9  
## Hedges' g(av) -0.7964846 0.497633 -1.684326 -0.06154947 0.9  
##   
## Note: SMD confidence intervals are an approximation. See vignette("SMD\_calcs").

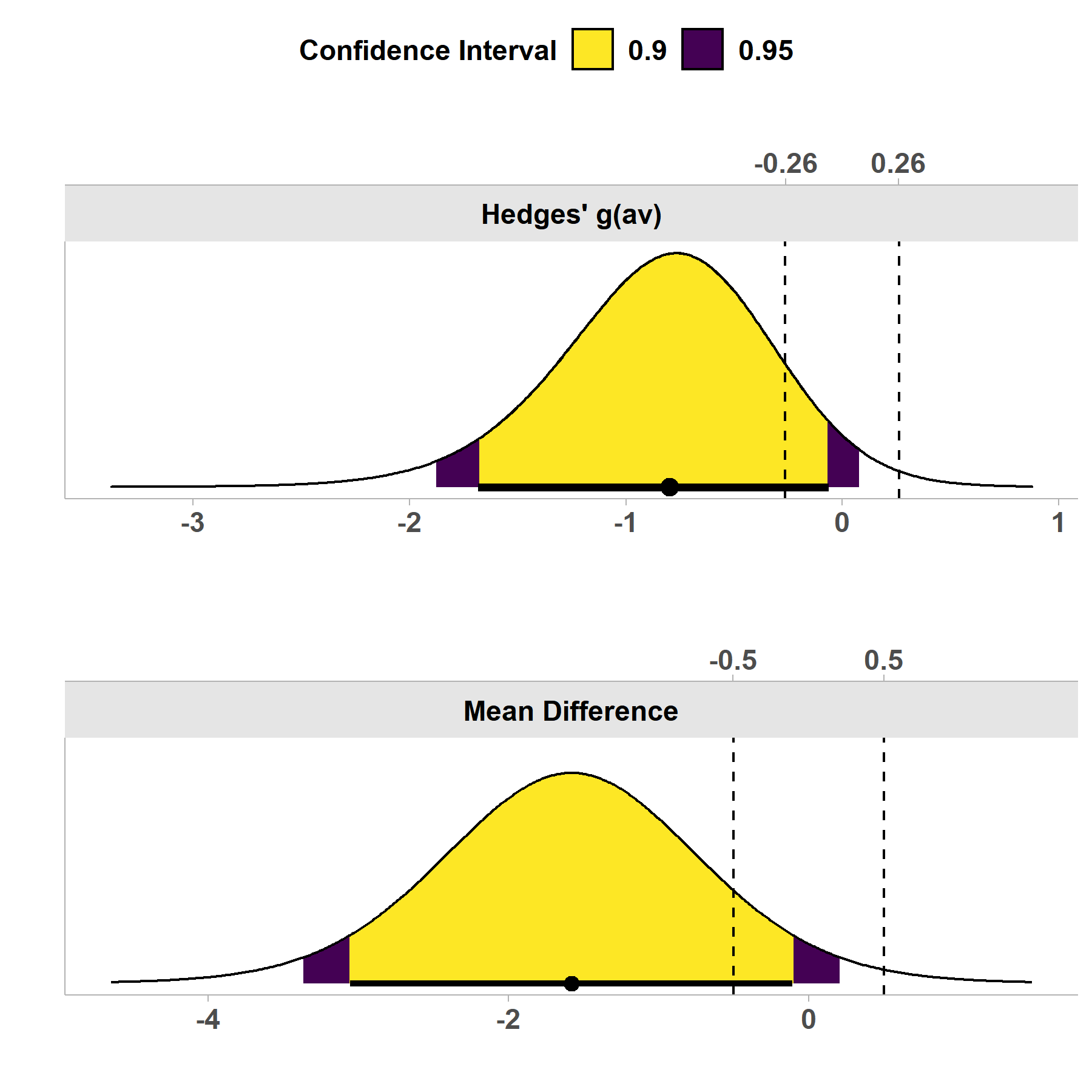
Another nice feature is the generic plot method that can provide a visual summary of the results. All of the plots in this package were inspired by the [concurve](https://cran.r-project.org/package=concurve) R package. There are two types of plots that can be produced. The first, and default, is the consonance density plot (type = "cd").

plot(res1, type = "cd")



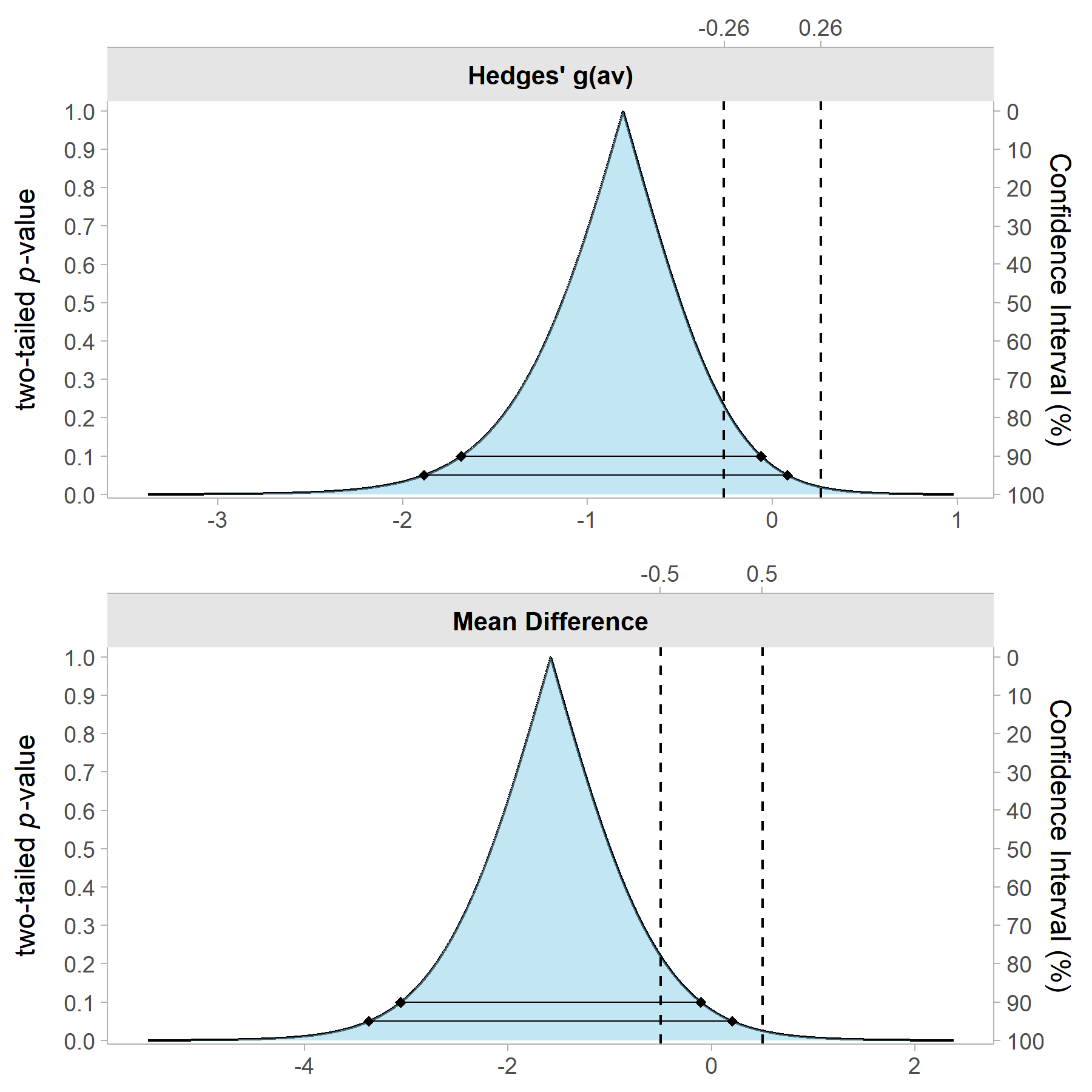
The shading pattern can be modified with the ci\_shades.

plot(res1, type = "cd",  
 ci\_shades = c(.9,.95))



Consonance plots, where all confidence intervals can be simultaneous plotted, can also be produced. The advantage here is multiple confidence interval lines can plotted at once.

plot(res1, type = "c",  
 ci\_lines = c(.9,.95))



# 4 Paired Samples

To perform a paired samples TOST, the process does not change much. We could process the test the same way by providing a formula. All we would need to then is change paired to TRUE.

res2 = t\_TOST(formula = extra ~ group,  
 data = sleep,  
 paired = TRUE,  
 low\_eqbound = -.5,  
 high\_eqbound = .5)  
res2

##   
## Paired t-test  
## Hypothesis Tested: Equivalence  
## Equivalence Bounds (raw):-0.500 & 0.500  
## Alpha Level:0.05  
## The equivalence test was non-significant, t(9) = -2.777, p = 9.89e-01  
## The null hypothesis test was significant, t(9) = -4.062, p = 2.83e-03  
## NHST: reject null significance hypothesis that the effect is equal to zero   
## TOST: don't reject null equivalence hypothesis  
##   
## TOST Results   
## t SE df p.value  
## t-test -4.062128 0.3889587 9 0.0028328902  
## TOST Lower -2.776644 0.3889587 9 0.9892407566  
## TOST Upper -5.347611 0.3889587 9 0.0002319027  
##   
## Effect Sizes   
## estimate SE lower.ci upper.ci conf.level  
## Raw -1.580000 0.3889587 -2.293005 -0.8669947 0.9  
## Hedges' g(z) -1.230152 0.2008070 -1.848296 -0.8362302 0.9  
##   
## Note: SMD confidence intervals are an approximation. See vignette("SMD\_calcs").

However, we may have two vectors of data that are paired. So instead we may want to just provide those separately rather than using a data set and setting the formula. This can be demonstrated with the “bugs” data.

res3 = t\_TOST(x = bugs$LDHF,  
 y = bugs$LDLF,  
 paired = TRUE,  
 low\_eqbound = -1,  
 high\_eqbound = 1)  
res3

##   
## Paired t-test  
## Hypothesis Tested: Equivalence  
## Equivalence Bounds (raw):-1.000 & 1.000  
## Alpha Level:0.05  
## The equivalence test was non-significant, t(90) = 2.655, p = 9.95e-01  
## The null hypothesis test was significant, t(90) = 6.649, p = 2.22e-09  
## NHST: reject null significance hypothesis that the effect is equal to zero   
## TOST: don't reject null equivalence hypothesis  
##   
## TOST Results   
## t SE df p.value  
## t-test 6.648618 0.2504032 90 2.223690e-09  
## TOST Lower 10.642177 0.2504032 90 6.681389e-18  
## TOST Upper 2.655058 0.2504032 90 9.953114e-01  
##   
## Effect Sizes   
## estimate SE lower.ci upper.ci conf.level  
## Raw 1.6648352 0.25040321 1.2486748 2.0809956 0.9  
## Hedges' g(z) 0.6940558 0.09548981 0.5296194 0.8707979 0.9  
##   
## Note: SMD confidence intervals are an approximation. See vignette("SMD\_calcs").

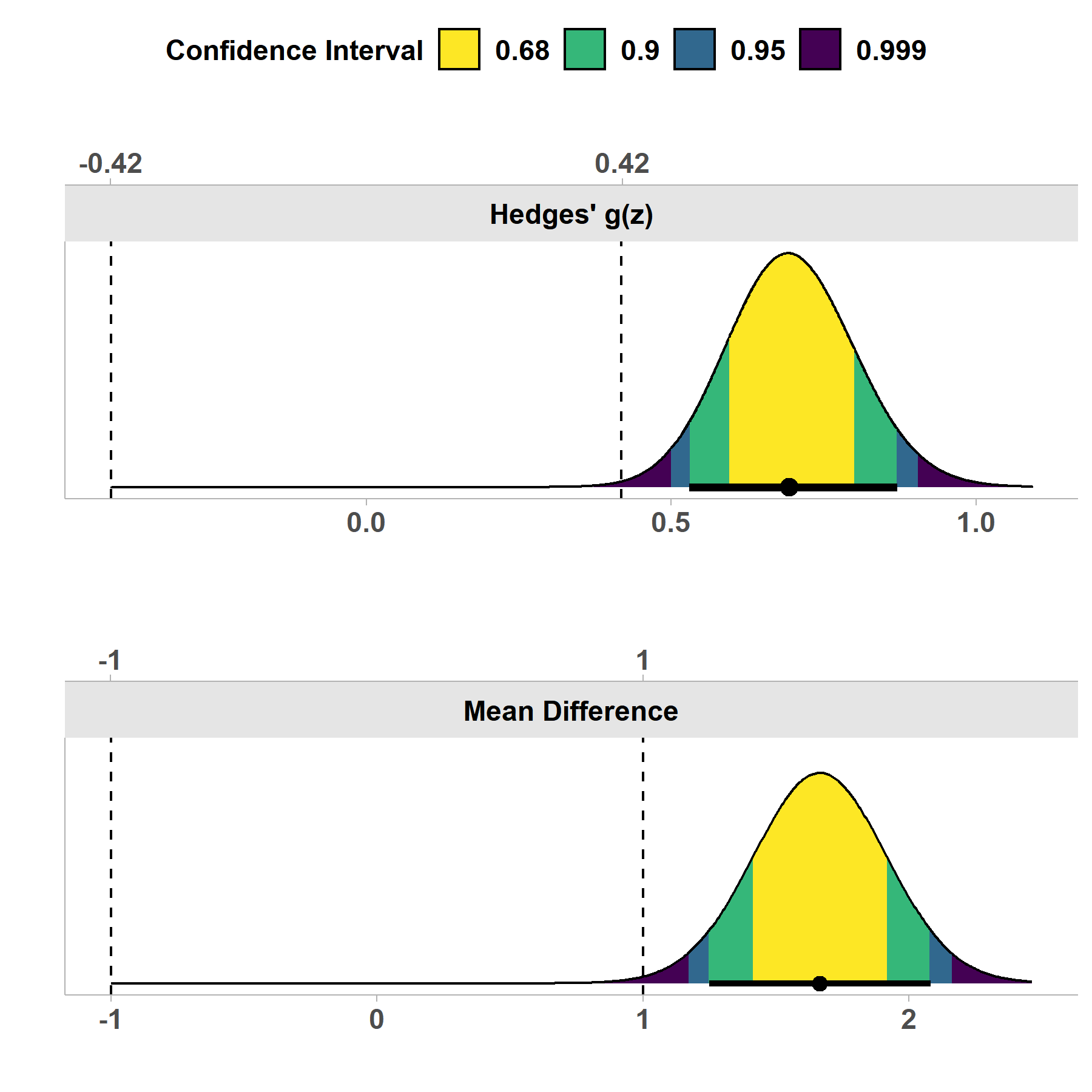
We may want to perform a Minimal Effect Test with the hypothesis argument set to “MET.”

res3a = t\_TOST(x = bugs$LDHF,  
 y = bugs$LDLF,  
 paired = TRUE,  
 hypothesis = "MET",  
 low\_eqbound = -1,  
 high\_eqbound = 1)  
res3a

##   
## Paired t-test  
## Hypothesis Tested: Minimal Effect  
## Equivalence Bounds (raw):-1.000 & 1.000  
## Alpha Level:0.05  
## The minimal effect test was significant, t(90) = 10.642, p = 4.69e-03  
## The null hypothesis test was significant, t(90) = 6.649, p = 2.22e-09  
## NHST: reject null significance hypothesis that the effect is equal to zero   
## TOST: reject null MET hypothesis  
##   
## TOST Results   
## t SE df p.value  
## t-test 6.648618 0.2504032 90 2.223690e-09  
## TOST Lower 10.642177 0.2504032 90 1.000000e+00  
## TOST Upper 2.655058 0.2504032 90 4.688649e-03  
##   
## Effect Sizes   
## estimate SE lower.ci upper.ci conf.level  
## Raw 1.6648352 0.25040321 1.2486748 2.0809956 0.9  
## Hedges' g(z) 0.6940558 0.09548981 0.5296194 0.8707979 0.9  
##   
## Note: SMD confidence intervals are an approximation. See vignette("SMD\_calcs").

Again, we can plot the effects from the t\_TOST result.

plot(res3a)



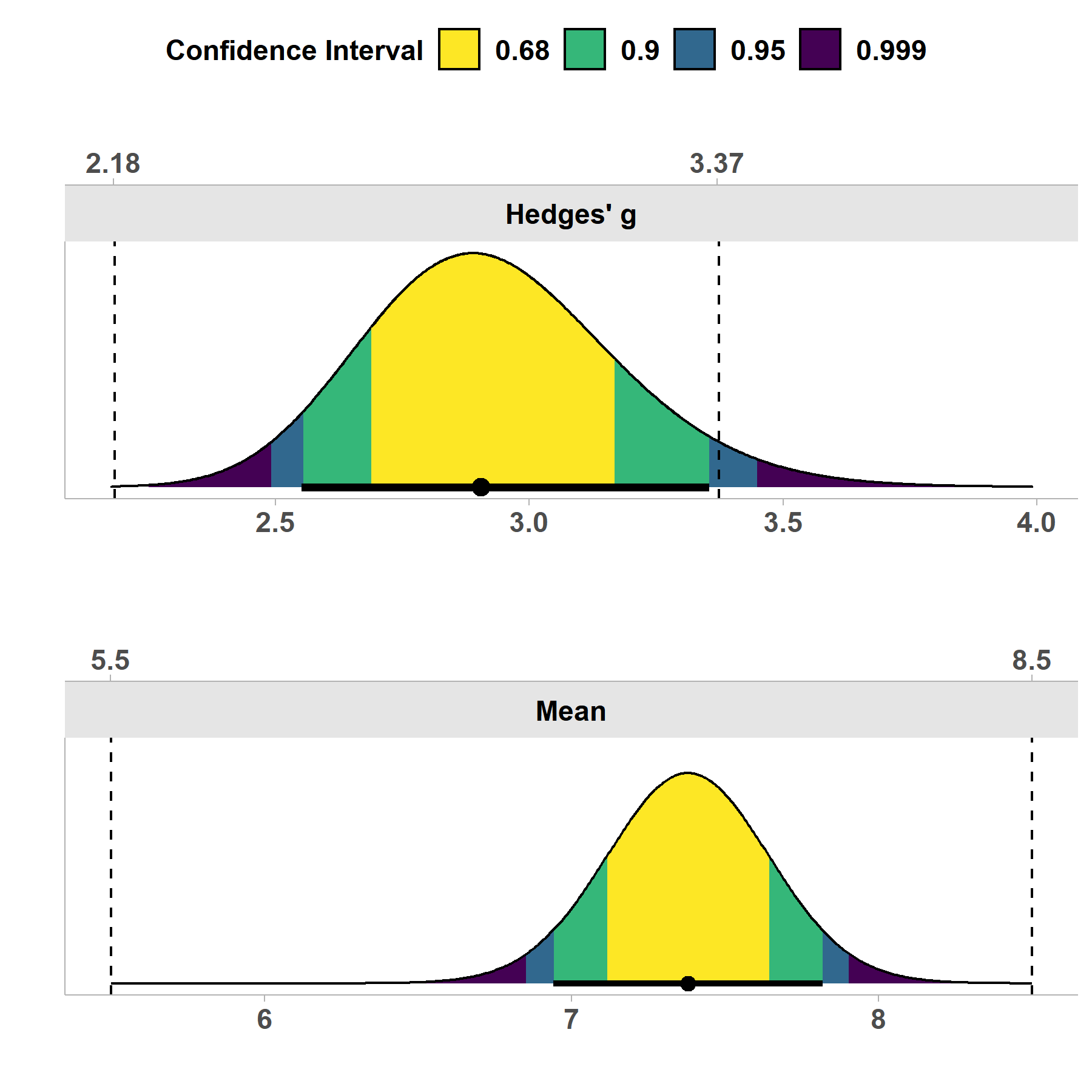
# 5 One Sample t-test

In other cases we may just have a one sample test. If that is the case all we have to do is supply the x argument for the data. For this test we may hypothesis that the mean of LDHF is not more than 1.5 points greater or less than 7.

res4 = t\_TOST(x = bugs$LDHF,  
 hypothesis = "EQU",  
 low\_eqbound = 5.5,  
 high\_eqbound = 8.5)  
res4

##   
## One Sample t-test  
## Hypothesis Tested: Equivalence  
## Equivalence Bounds (raw):5.500 & 8.500  
## Alpha Level:0.05  
## The equivalence test was significant, t(90) = -4.244, p = 2.66e-05  
## The null hypothesis test was significant, t(90) = 27.942, p = 3.91e-46  
## NHST: reject null significance hypothesis that the effect is equal to zero   
## TOST: reject null equivalence hypothesis  
##   
## TOST Results   
## t SE df p.value  
## t-test 27.942403 0.2640833 90 3.906969e-46  
## TOST Lower 7.115638 0.2640833 90 1.298575e-10  
## TOST Upper -4.244416 0.2640833 90 2.658068e-05  
##   
## Effect Sizes   
## estimate SE lower.ci upper.ci conf.level  
## Raw 7.379121 0.2640833 6.940225 7.818017 0.9  
## Hedges' g 2.904671 0.2452067 2.551811 3.354884 0.9  
##   
## Note: SMD confidence intervals are an approximation. See vignette("SMD\_calcs").

plot(res4)



# 6 Only have the summary statistics? No problem!

In some cases you may only have access to the summary statistics. Therefore, we created a function, tsum\_TOST,to perform the same tests just based on the summary statistics. This involves providing the function with a number of different arguments.

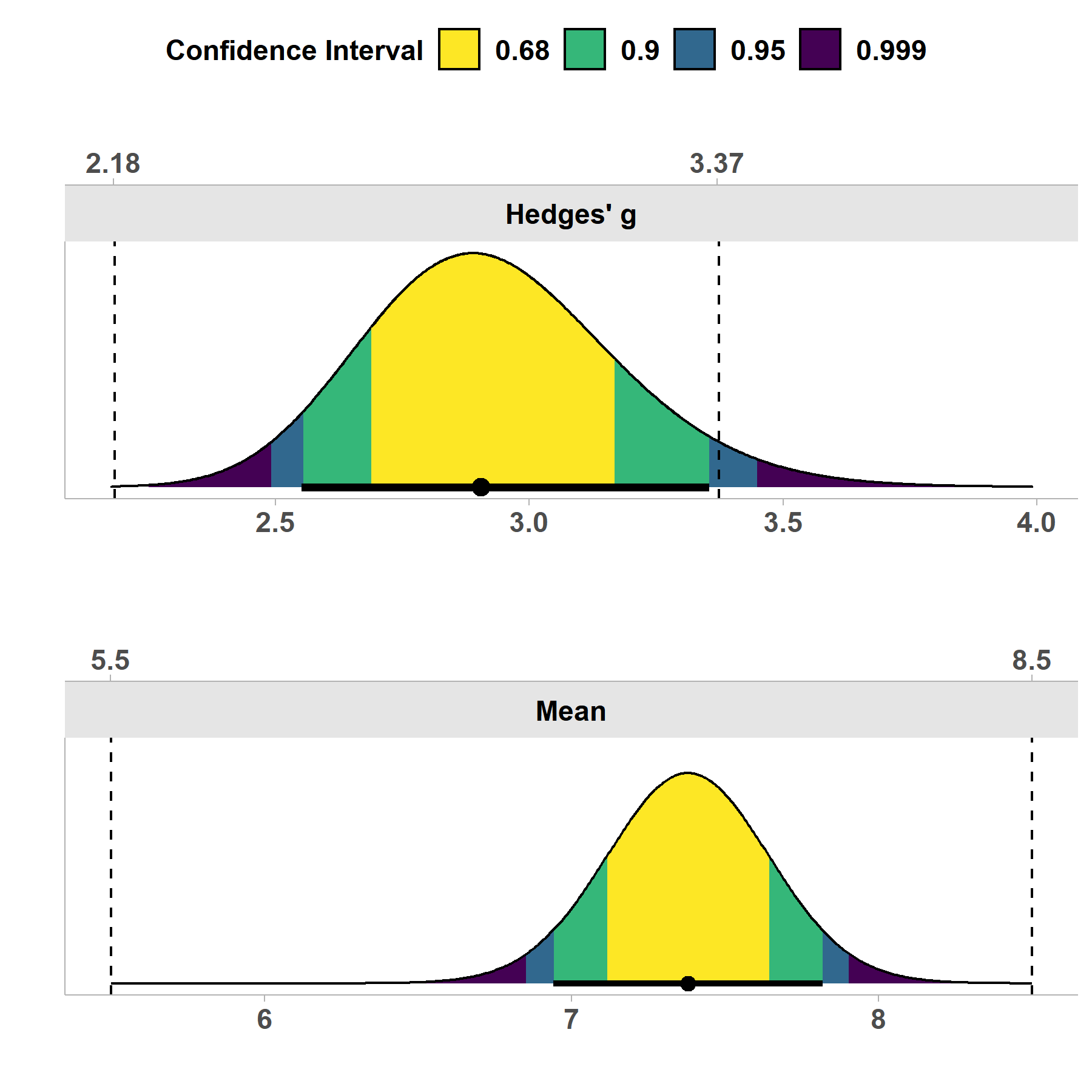
* n1 & n2 the sample sizes (only n1 needs to be provided for one sample case)
* m1 & m2 the sample means
* sd1 & sd2 the sample standard deviation
* r12 the correlation between each if paired is set to TRUE

The results from above can be replicated with the tsum\_TOST

res\_tsum = tsum\_TOST(  
 m1 = mean(bugs$LDHF, na.rm=TRUE),  
 sd1 = sd(bugs$LDHF, na.rm=TRUE),  
 n1 = length(na.omit(bugs$LDHF)),  
 hypothesis = "EQU",  
 low\_eqbound = 5.5,  
 high\_eqbound = 8.5  
)  
  
res\_tsum

##   
## One-sample t-Test  
## Hypothesis Tested: Equivalence  
## Equivalence Bounds (raw):5.500 & 8.500  
## Alpha Level:0.05  
## The equivalence test was significant, t(90) = -4.244, p = 2.66e-05  
## The null hypothesis test was significant, t(90) = 27.942, p = 3.91e-46  
## NHST: reject null significance hypothesis that the effect is equal to zero   
## TOST: reject null equivalence hypothesis  
##   
## TOST Results   
## t SE df p.value  
## t-test 27.942403 0.2640833 90 3.906969e-46  
## TOST Lower 7.115638 0.2640833 90 1.298575e-10  
## TOST Upper -4.244416 0.2640833 90 2.658068e-05  
##   
## Effect Sizes   
## estimate SE lower.ci upper.ci conf.level  
## Raw 7.379121 0.2640833 6.940225 7.818017 0.9  
## Hedges' g 2.904671 0.2452067 2.551811 3.354884 0.9  
##   
## Note: SMD confidence intervals are an approximation. See vignette("SMD\_calcs").

plot(res\_tsum)



# 7 Power Analysis for t-test based TOST

A new function, power\_t\_TOST has been created to allow for power calculations for TOST analyses that utilize t-tests. This function uses a more accurate method than the older functions in this package and match the results of the PASS statistics software. The exact calculations of power are based on Owenâ€™s Q-function or by direct integration of the bivariate non-central t-distribution (inspired by (**PowerTOST?**) in the PowerTOST R package). Approximate power is implemented via the non-central t-distribution or the â€˜shiftedâ€™ central t-distribution (**diletti1991?**). The function is limited to power analyses involves one sample, two sample, and paired sample cases. More options are available in the PowerTOST R package.

The interface for this function is quite simple and was intended to mimic the base R function power.t.test. The user must specify the 2 equivalence bounds, and leave only one of the other options blank (alpha, power, or n). The “true difference” can be set with delta and the standard deviation (default is 1) can be set with the sd argument. Once everything is set and the function is run, the function will return a object of the power.htest class.

As an example, let’s say we are looking at an equivalence study where we assume the true difference is 1 unit, the standard deviation is 2.5, and we set the equivalence bounds to 2.5 units as well. If we want to find the sample size adequate to have 95% power at an alpha of 0.025 we enter the following:

power\_t\_TOST(n = NULL,  
 delta = 1,  
 sd = 2.5,  
 low\_eqbound = -2.5,  
 high\_eqbound = 2.5,  
 alpha = .025,  
 power = .95,  
 type = "two.sample")

##   
## Two-sample TOST power calculation   
##   
## power = 0.95  
## beta = 0.05  
## alpha = 0.025  
## n = 73.16747  
## delta = 1  
## sd = 2.5  
## bounds = -2.5, 2.5  
##   
## NOTE: n is number in \*each\* group

From the analysis above we would conclude that adequate power is achieved with 74 participants per group and 148 participants in total.

# 8 Non-Parametric Tools

Why this may be useful Describe these arenâ€™t tests of medians but symmetry tests (myth busting) Paired samples may want to be rank transformed

# 9 Bootstrapping

Detail the boostrapping tools

# 10 Equivalence Testing with ANOVAs

Detail the tools from Campbell & Lakens integrated with this package.

# Acknowledgement(s)

An unnumbered section, e.g. \section\*{Acknowledgements}, may be used for thanks, etc. if required and included *in the non-anonymous version* before any Notes or References.

# Disclosure statement

An unnumbered section, e.g. \section\*{Disclosure statement}, may be used to declare any potential conflict of interest and included *in the non-anonymous version* before any Notes or References, after any Acknowledgements and before any Funding information.

# Funding

An unnumbered section, e.g. \section\*{Funding}, may be used for grant details, etc. if required and included *in the non-anonymous version* before any Notes or References.

# Notes on contributor(s)

An unnumbered section, e.g. \section\*{Notes on contributors}, may be included *in the non-anonymous version* if required. A photograph may be added if requested.

# Nomenclature/Notation

An unnumbered section, e.g. \section\*{Nomenclature} (or \section\*{Notation}), may be included if required, before any Notes or References.

# Notes

An unnumbered Notes section may be included before the References (if using the endnotes package, use the command \theendnotes where the notes are to appear, instead of creating a \section\*).

# 11 References

## 11.1 References cited in the text

## 11.2 The list of references

References should be listed at the end of the main text in alphabetical order by authors’ surnames, then chronologically (earliest first). If references have the same author(s), editor(s), etc., arrange by year of publication, with undated works at the end. A single-author entry precedes a multi-author entry that begins with the same name. If the reference list contains two or more items by the same author(s) in the same year, add a, b, etc. and list them alphabetically by title. Successive entries by two or more authors when only the first author is the same are alphabetized by co-authors’ surnames. If a reference has more than ten named authors, list only the first seven, followed by ‘et al.’ If a reference has no author or editor, order by title; if a date of publication is impossible to find, use `n.d.’ in its place.

The following list shows some sample references prepared in the Taylor & Francis Chicago author-date style.

(Adelman, 2009; Albiston, 2005)

Appendix

Any appendices should be placed after the list of references, beginning with the command followed by the command for each appendix title, e.g.

produces:

Subsections, equations, figures, tables, etc. within appendices will then be automatically numbered as appropriate. Some theorem-like environments may need to have their counters reset manually (e.g. if they are not numbered within sections in the main text). You can achieve this by using (for example) just after the command.

Please note that if the package is used on a document containing appendices, the command must be included immediately before the command in order to ensure that the floats in the main body of the text are numbered as such.

%%%% See above for an explanation of why this command might be needed.

Authors may occasionally encounter problems with the preparation of a manuscript using . The appropriate action to take will depend on the nature of the problem:

This article template and the class file may be obtained via the `Instructions for Authors’ pages of selected Taylor & Francis journals.

Please note that the class file calls up the open-source  packages booktabs.sty, epsfig.sty and rotating.sty, which will, for convenience, unpack with the downloaded template and class file. The template calls for natbib.sty and subfigure.sty, which are also supplied for convenience.

This article template, the class file and the associated open-source  packages are also available via e-mail. Requests should be addressed to , clearly stating for which journal you require the template and class file.

Adelman, R. (2009). ‘Such stuff as dreams are made on’: God’s Footstool in the Aramaic Targumim and Midrashic tradition. New Orleans, Louisiana: Paper presented at the annual meeting for the Society of Biblical Literature.

Albiston, C. R. (2005). Bargaining in the shadow of social institutions: Competing discourses and social change in the workplace mobilization of civil rights. *Law and Society Review*, *39*(1), 11–47.