neg.indvars Function (negligible package)

Rob Cribbie

2023-09-13

The neg.indvars function exists within the *negligible* R package. The negligible package is available on both CRAN and github (https://github.com/cribbie/negligible).

The purpose of the neg.indvars function is to evaluate if multiple population variances can be considered *negligibly* different (i.e., practically equivalent). For example, imagine that you are running a two independent samples *t* test and you want to test the assumption that the population variances are equal. You could use the neg.indvars to evaluate that assumption.

The null hypothesis specifies that the difference in the variances is non-negligible, while the alternate hypothesis states that the difference in the variances is negligible. See Kim and Cribbie (2018) for more details.


The basic set-up of the function looks like this: neg.indvars(dv, iv, eps = 0.5, alpha = 0.05, na.rm = TRUE, data = NULL)

There are three arguments that are required (i.e., do not have defaults): dv - dependent/outcome variable iv - independent/predictor variable (must be a factor) data - name of the dataset where the dv and iv reside

There are also other optional arguments: eps - Stefan Wellek suggests conservative (eps = .25) and liberal (eps = .50) bounds for the test of negligible difference in independent population variances. The default is .50, but any value could be used. See Mara & Cribbie (2017) or Wellek (2010).

alpha - nominal Type I error rate. The default is .05, but any value can be used (e.g., .01, .10, .06)

na.rm - should cases with missing values be deleted. Right now the function only works with na.rm=TRUE if there are missing values.
d <- mtcars # open the dataset and store it in the object d
names(d) # look at the variable names

Let’s look at an example using the mtcars dataset from R. Our outcome variable is miles/gallon (mpg) and our predictor is number of cylinders in the vehicle (cyl). In this example, we want to know if the difference in the population variances of mpg across cyl (4, 6, 8) is negligible (i.e., are the population variances of mpg across cyl equivalent).

```r
## [1] "mpg" "cyl" "disp" "hp" "drat" "wt" "qsec" "vs" "am" "gear"
## [11] "carb"
```

d$cyl <- factor(d$cyl)
tapply(d$mpg, d$cyl, var) # explore the variances of mpg at each level of cyl

```r
##          4          6          8
## 20.338545  2.112857  6.553846
```

Now, let’s apply the neg.vars function to see if we can reject the null hypothesis that the population variances are non-negligibly different.

```r
library(negligible)

## Warning: package 'negligible' was built under R version 4.3.2

neg.indvars(dv = mpg, iv = cyl, eps = .5, alpha = .05, na.rm = TRUE, data = d)

## -- Equivalence of Population Variances --
## -- Independent Groups --
##
## Group Variances:
##          4          6          8
## 20.338550  2.112857  6.553846
##
## Group Standard Deviations:
##  4.509828  1.453567  2.560048
##
## Group Median Absolute Deviations:
##  6.52344  1.92738  1.55673
##
## *******************
##
## Ratio of Largest to Smallest Variances:
##  9.626086
##
## *******************
##
## Epsilon Value (establish the Equivalence Interval):
##  0.5
```
Levene-Wellek-Welch (LWW) Statistic:
1.147718

Critical Value for LWW:
0.03458143

NHST Decision:
The null hypothesis that the differences between the population variances falls outside the equivalence interval cannot be rejected. Therefore, we cannot conclude that the population variances cannot be concluded. Be sure to interpret the magnitude (and precision) of the effect size.

The LWW statistic must be less than the critical value to reject the null hypothesis. In this case, LWW = 1.148 > LWW(crit) = .035, so we cannot reject Ho. There is a lack of support for the contention that the population variances are negligibly different.