**Worst Case Resistance Testing: A Nonresponse Bias Solution for Today’s Survey Research Realities - Supplemental Material**

This supplemental material contains the two appendices for the “Worst Case Resistance Testing: A Nonresponse Bias Solution for Today’s Survey Research Realities” paper. Appendix A contains optimization algorithms for the single sample t test and correlation test WRCT methods. Appendix B contains inference and an optimization algorithm for the two-sample test WRCT methods.

APPENDIX A: Optimization Algorithms

Optimization Algorithm for Single Sample t Test Inference

Equation (9) in the main paper cannot be solved outright as both and are dependent on *n*2, creating cross-dependencies. However, the equation can be solved using a simple fixed-point algorithm. First rearrange (9) as follows, for the critical point, setting .

|  |  |
| --- | --- |
|  | (A-) |

1. Utilize an initial starting value of *n*2 = *n*1and call this *nOpt*.
2. Calculate ,, and *s*c using *nOpt*.
3. Calculate *n*2 from equation (9), using *n*2 = *nOpt* in the RHS of the equation, and store this in variable *nCalc*, i.e., .
4. Recalculate *nOpt* as (*nOpt*+*nCalc*)/2.
5. Repeat steps 2-4 until |*nOpt*-*nCalc*|<*δ*, where *δ* is some pre-set convergence criterion.

In practice, the values of *nOpt* and *nCalc* converge so that |*nOpt*-*nCalc*|<*δ*.

Optimization Algorithm for Correlation Inference

The fixed-point method used for the one-sample tests did not converge for the correlation test, due to Equation (16) in the main paper having both a negative and positive root. Thus, a divide and conquer optimization method was employed. It takes advantage of the fact that given a candidate value of *zrc*, the value of *n*2 can be calculated by rearranging Equation (12) in the main paper as follows:

|  |  |
| --- | --- |
|  | (A-) |

Collecting *n*2 terms gives (A-3).

|  |  |
| --- | --- |
|  | (A-) |

Rearranging in terms of *n*2 gives (A-4).

|  |  |
| --- | --- |
|  | (A-) |

1. The algorithm works by exploring the possible values of *zr*c, calculating *z* and then constraining z towards . Calculate *zr*1 from *r*1. Calculate *zr*2 from *r*2. For a nonresponse effect size *r*2, the steps are as follows:
2. From (12) in the main paper, *zr*c is a linear combination of *zr*1 and *zr*2, so lies between these two values. For scenarios 1 and 4 in Table 2 in the main paper, set LB = *zr*2 and UB = *zr*1*;* for scenarios 2 and 3 set LB = *zr*1 and UB = *zr*2*.*
3. Set  and then use this value of to calculate *n*2, using (A-4).
4. Calculate the value of .
5. Now, if , set *LB* =  else set *UB* = .
6. If , where δ is a convergence criterion then exit. Otherwise go to step 3.

If the value of is not in the range of the initial [LB,UB] this indicates that there is no possible *n*2 for the selected *r*2 that can give a z that reaches the critical value. Typically, for scenarios 1 and 4, when increasing effect size*,* this occurs when *r2* is around 0, i.e., where the trade-off of the low effect size against high *n*2reaches an equilibrium.

APPENDIX B: Two-Sample Tests

The calculations of the number of items required to change significance for two-sample independent sample tests work in a similar manner to the calculations for the single sample tests[[1]](#footnote-1). Consider data sampled from two distinct populations A and B and the following definitions:

*d*0: The hypothesized difference between the population means.

: The sample means.

*n1A*, *n1B*: The sizes of the samples.

*s*1*A*, *s*1*B*: The sample standard deviations.

Three possible tests are outlined below.

A z test:

 (B-1)

The Welch’s t test:

, as per (10), where  (B-2)

The Student’s t test:

, where.  (B-3)

In addition, the pooled standard deviation can be defined as,

 (B-4)

Here we wish to find the minimum *n*2*A*, *n*2*B*, so that the direction of the hypothesis is reversed. There are several differences between a single sample test and a two-sample test. First, the fact that both groups can have different value of *n* introduces additional degrees of freedom. In fact, one could find infinite solutions to the problem by repeatedly increasing *n2A* and decreasing *n2B*. To get around this, a single multiplier *ϕ* is used, so that  and .

There are several measures of effect size for two-sample tests. Hedge’s g utilizes the pooled standard deviation and Cohen’s d gives the maximum likelihood estimate for the pooled standard deviation (McGrath and Meyer, 2006).

 (B-5)

 (B-6)

For a consistent definition of pooled standard deviation, we use (B-5).

For a right tailed significance test where *t* *>* *tcrit[[2]](#footnote-2)*, consider additional data sampled from the population with some overall effect size *g2* that is less than *g*1, such that there is some finite *φ*, so that  and , and the test on a combined samples of  and  will not be significant, i.e. *t* *≤* *tcrit*.

The following variables are defined for the appended data:

*g*2: The effect size for the appended data.

: The sample means.

*n*2*A*, *n*2*B*: The sizes of the samples.

*s*2*A*, *s*2*B*: The sample standard deviations for the appended data. As before, one can define some *θ* , where  and  to create bounds for the WRCT *n* values. However, for the remainder of the derivation, the simplifying assumption is made that *θ* = 0 and that the group variances for the appended data are the same as for the sample data.

The effect size for the second sample is given below.

 (B-7)

This can be rewritten as follows:

 (B-8)

The effect size gives a difference between the groups, but not the exact location. To ensure feasibility, we make the simplifying assumption of keeping the center of the groups to be the same, i.e.,

 (B-9)

Thus:

 (B-10)

 (B-11)

When it comes to weighting by *n*, we use the relative weights (1 for the original data and φ for the appended data). The aggregate value for the difference between the two means for the combined sample can be given as follows:

  (B-12)

The associated value for the standard deviation for the combined data for sample A is given below, adapted from the combined standard deviation formula in Higgins (2019). A similar equation can be given for sample B.

 (B-13)

The value of *φ* needs to be calculated to derive the WRCT sample sizes  and . For a right-sided test where we wish to find the lowest n for which *t* *≤* *tcrit*, define some small quantity ε, such that .

 (B-14)

The same calculation can be used for Welch’s t test, but with the degrees of freedom given in (B-15).

 (B-15)

For the student t test:

, with  (B-16)

As per the one-sample test and Table 2 in the main paper, there are four combinations of left-sided/right sided and significant/non-significant tests. A local search algorithm is utilized to calculate the value of value of *φ* needed to reverse the test for a given WCRT effect size *g2*. The algorithm steps are given below.

1. Utilize an initial starting value of *φ =* 1*.* Call this variable *φOpt* and use it to calculate *n*2*A*and n2*B*. Also define an algorithm step size *φChange* (with initial default at *φChange =* 0.5), *MaxSameDirection* to help control the speed/sensitivity of the search, and a convergence criterion *δ* for the test statistic (default at 1E-6).
2. Calculate , , , , *s*c*A*, *s*c*B*, and for the Student-t test, *scp*.
3. Calculate the value of the test statistic (t or z), which we will denote *tz*. If the test statistic is sufficiently close to the critical value so that then terminate the algorithm.
4. For scenarios one and four in Table 2, increase the value of *φOpt* by *φChange* if and decrease the value of *φOpt* by *φChange* if . The opposite action should be taken for scenarios two and three.
5. Update *φChange* to control algorithm convergence*.* If the move direction has changed from the previous move (e.g., increase to decrease) then divide *φChange* by2*.* If *MaxSameDirection* consecutive moves have occurred in one direction, then multiply *φChange* by2.
6. Repeat steps 2 to 5 until algorithm convergence is found in step 3.

**References**

McGrath, R. E., & Meyer, G. J. (2006). When effect sizes disagree: the case of r and d. *Psychological Methods*, *11*(4), 386.

1. We do not include paired sample tests, as these can be implemented as a single sample test on the difference between groups [↑](#footnote-ref-1)
2. The same logic applies to the two-sample independent samples z test. [↑](#footnote-ref-2)