The SSER Method: Replicability Possibilities within the General Linear Model Extended to the Independent Samples *t*-Test, One-Way ANOVA, and Chi-Square

David Walker
Northern Illinois University

The purpose of this examination was to extend the research pertaining to the idea of a statistically significant exact replication (SSER) method for estimating a study's replicability for the cases of the independent sample t-test, the one-way analysis of variance, and chi-square. A second intention of this study was to provide users with three programs that would calculate the SSER value when there was a statistically significant finding to assist in determining the chance that an exact replication would be statistically significant beyond 50%.

Over a 10 year period of applied and theoretical research pertaining to the statistically significant exact replication (SSER) technique, and other concepts affiliated with the SSER such as replication, power, and probability, the literature indicated that developmental work and scholarly debate in this area have resulted in an probability-based method for estimating a study's replicability (cf. Froman & Shneyderman, 2004; Greenwald, Gonzalez, Harris, & Guthrie, 1996; Macdonald, 2002, 2003; Newman, McNeil, & Fraas, 2004; Posavac, 2002, 2003; Walker, 2006). The SSER is premised on the idea that, "… the probability of a statistically significant exact replication (SSER) can be estimated from the probability of the statistical test" (Newman et al., p. 37). Within the idiom of the SSER, the concept of replication was operationalized as "…a test conducted with additional subjects sampled in the same fashion as those in the initial study and tested under conditions identical to those of the initial study" (Greenwald et al., p. 181). That is, the "…initial experiment and the replication differ only due to random variation (Posavac, 2002, p. 102). Because SSER probability is an estimate derived from a probability value of an observed test statistic that an exact replication will be statistically significant, it should be thought of as an upper bound value of replicability ranging between 0 and 1.00, with a benchmark of ≥ .80 to assure a demonstrable result of the likelihood of a finding being repeated successfully (Greenwald et al.). Ultimately, the SSER method poses the subsequent question: How much beyond a 50% chance is there of replicating a statistically significant finding for an observed sample statistic?

Purpose

This study had two purposes. The first intention was to expand on and add to the aforementioned scholarly literature in this area of research. The second goal was to provide users of the method with three programs in SPSS (Statistical Package for the Social Sciences) that would calculate the SSER value when there was a statistically significant finding extended to three cases all algebraically related: the independent samples t-test, the one-way analysis of variance (ANOVA), and chi-square (cf. Walker, 2006 for previous extension work with the t-test and ANOVA). In statistics, the issue of *power and sample size* is related to the Type I error rate, alpha level of significance, directional nature of the hypothesis, and population variance.

Method

Using the data provided in Newman et al. (2004) for the case of an independent samples t-test with an observed t value of 2.150, 38 df (degrees of freedom), and a two-tailed t critical value of 2.024 at the .05 alpha level, the concept of the SSER method was replicated and modified with the t-test example and extended to the one-way ANOVA and chi-square. For the t-test program, users need to supply within the program's syntax matrix the observed t value and the test's df, which is ascertained by n-2. For the ANOVA, users need to provide the observed F value, df1 for the numerator, df2 for the denominator, df3 for the R2 effect size, where df3 is determined via n-1, and the sample size. For the chi-square program, users need to supply the observed chi-square value, the df, and the sample size (see Appendices A – C for the programs' syntax). Once these sample-based, observed data are entered in the marked area of a particular program's syntax, users run the program to derive an SSER result. The programs' defaults are set for all of the critical values at the .05 level. If this default level needs to be changed a priori, due to theoretical assumptions and/or literature-based reasoning, users can follow the instructions embedded within the programs' syntax.

Table 1. SSER results

Test	*df*	Observed Statistic Value	Critical Value (.05)	*p*-Value	Effect Size	SSER Value
t-Test	38	2.150	2.024	.038	.698	.550
ANOVA	1,38	4.623	4.098	.038	.106	.527
Chi-Square	1	4.320	3.841	.038	.312	.511

Results and Discussion

To answer the SSER's question, how much beyond a 50/50 chance is there of replicating a statistically significant finding for an observed sample statistic?, the Newman et al. (2004) example calculated an SSER value for the *t*-test of .55 or just a little over half of the replications would be anticipated to generate an observed *t* value greater than 2.150 and a little less than half of the replications would be expected to yield an observed *t* value less than 2.150. In replication of said result, and extending the method to two other tests, data in Table 1 indicated that for all three of the statistically significant test results, there was just a slight likelihood of over a 50% chance of replication, with the upper bound of an exact statistically significant replication estimated at .55 for the *t*-test, .53 for the ANOVA, and .51 for chi-square; all of which were not very reliable. In addition, if we report the effect sizes affiliated with these statistically significant results that had just over a 50% chance of replication, we find that the *t*-test had a Cohen's *d* = 0.70, the ANOVA had an $R^2 = 0.11$, and chi-square had a coefficient of contingency = .31. Even though we had statistically significant results and effects sizes that ranged from small to bordering on large, the important information garnered from these data are that they had just over a 50% chance of replication.

In a second example using the ANOVA program, the probability value of the observed F was statistically significant at *p* = .049, but the SSER value was .001 or virtually no chance, beyond 50%, that an exact replication would be statistically significant at the .05 level. This outcome illustrates a caution noted by Newman et al. (2004) that a statistically significant result affiliated with an observed test value will not always generate an exact replication that will be statistically significant as well.

It should be noted that because a given SSER value is based on a sample test statistic, which in turn is related to a sample size, the 50/50 split of an exact replication being statistically significant beyond 50% is assumed via a normal distribution and predicated on the fact that "larger sample sizes… would give the researcher more statistical power and would increase the likelihood of rejecting null hypotheses in SSERs" (Posavac, 2002, p. 111). Given these assumptions, there is a possibility that an SSER value could be lower, in the sense of not reaching its maximum upper value, when affiliated with a small sample size (i.e., *n* < 30). Posavac (2002) determined that small samples had limited impact on SSER probabilities by showing, in the case of the *t*-test, that when a sample was as small as *n* = 10 or df = 8 with alpha established at the levels of .05, .01, and .005 for a two-tailed test, values for the SSER were .50 at the .05 level, ranging from .73 to .84 at the .01 level, and between .80 to .92 at the .005 level.

Although the SSER is a post-hoc viewpoint, it does not carry the same connotation as selective post-hoc analyses for data dredging since it would only be performed after a statistically significant result were found and gives one an indication if said result were replicable or not beyond 50%. Also, it should be emphasized that the terms of alpha level setting (i.e., to control against type I error) need to be determined by users of the programs *a priori* and based on theory and/or research, and that these programs, while undemanding to run, should only be employed when statistical significance has been realized. Finally, the SSER replication probability, like other probability indices, should be interpreted in its research context with attentiveness toward factors that may impact its value such as influential data points, sampling variability, or data distribution (Macdonald, 2002).

Conclusion

The purpose of this study was to extend the research pertaining to the idea of a statistically significant exact replication method. The current study continues the idea of the SSER method and provides users with programs that will calculate the SSER value when there is a statistically significant finding to assist in determining the chance that an exact replication will be statistically significant beyond 50%. A standard feature of science is replication and an extension of the SSER method within the general linear model should afford users with more data upon which to base their decisions pertaining to, for example, the reliability of particular variables in a model or result stability.

References

Froman, T., & Shneyderman, A. (2004). Replicability reconsidered: An excessive rangeof possibilities. *Understanding Statistics, 3*(4), 365-373.

Greenwald, A. G., Gonzalez, R., Harris, R. J., & Guthrie, D. (1996). Effect sizes and *p*-values: What should be the reported and what should be replicated? *Psychophysiology, 33*(1), 175-183.

Macdonald, R. R. (2002). The incompleteness of probability models and the resultant implications for theories of statistical inference. *Understanding Statistics, 1*(3), 167-189.

Macdonald, R. R. (2003). On determining replication probabilities: Comments on Posavac (2002). *Understanding Statistics, 2*(1), 69-70.

Newman, I., McNeil, K., & Fraas, J. (2004). Two methods of estimating a study's replicability. *Mid-Western Educational Researcher, 17*(2), 36-40.

Posavac, E. J. (2002). Using p values to estimate the probability of a statistically significant replications. *Understanding Statistics, 1*(2), 101-112.

Posavac, E. J. (2003). Response to Macdonald. *Understanding Statistics, 2*(1), 71-72.

Walker, D. A. (2006). The statistically significant exact replication method: A programming and conceptual extension. *Mid-Western Educational Researche*r, *19*(4), 7-11.

Send correspondence to: David Walker
Northern Illinois University
Email: dawalker@niu.edu

Appendix A: SSER program for the independent samples *t*-test

```
*****************************************************************************
NOTE: During your initial research, if the probability value of the observed t-test is > .05, there
is NO need to run this program
*****************************************************************************.
DATA LIST LIST / TOBS (F9.3) DF (F8.0).
*****************************************************************************
Between BEGIN DATA and END DATA below, put your observed t value (TOBS) and degrees of
freedom (DF, which is n-2)
*****************************************************************************.
BEGIN DATA
2.150    38
END DATA.
*****************************************************************************
NOTE: Below in TDIFF for the critical value of t, choose the alpha level for a two-tailed test, either
TCRIT.05, TCRIT.01, or TCRIT.001 Currently, the program default is set at the .05 level
*****************************************************************************.
COMPUTE TCRIT.05 = ABS(IDF.T(.025,DF)).
COMPUTE TCRIT.01 = ABS(IDF.T(.005,DF)).
COMPUTE TCRIT.001 = ABS(IDF.T(.0005,DF)).
COMPUTE TDIFF = TOBS-TCRIT.05.
COMPUTE TREP = CDF.T(TDIFF,DF).
COMPUTE SIG1 = CDF.T(TOBS,DF).
COMPUTE D = 2*TOBS/SQRT(DF).
COMPUTE SIG = (1-SIG1)*2.
EXECUTE.
FORMAT TCRIT.05 TO SIG  (F9.3).
VARIABLE LABELS D 'Cohens d Effect Size (.20, .50, .80 are Suggested = Small, Medium, and Large
Effects)'/TOBS 'Your Observed t Value'/SIG 'The Probability of Your Observed t Value'/TCRIT.05 'For
Your DF, the Critical Value of t, Alpha=.05, Two-Tailed Test (Program Default Value)'/TCRIT.01 'For
Your DF, the Critical Value of t, Alpha=.01, Two-Tailed Test'/TCRIT.001 'For Your DF, the Critical
Value of t, Alpha=.001, Two-Tailed Test'/DF 'Degrees of Freedom'/TREP 'The Upper Limit of the SSER
Probability Value'/.
REPORT FORMAT=LIST AUTOMATIC ALIGN (CENTER)
  /VARIABLES= DF TOBS TCRIT.05 SIG D TCRIT.01 TCRIT.001
  /TITLE "Test Statistics".
REPORT FORMAT=LIST AUTOMATIC ALIGN (LEFT)
MARGINS (*,120)
  /VARIABLES= TREP
  /TITLE "How Much Beyond a 50/50 Chance Do You Have of Replicating Your Statistically Significant
Findings?".
************************************************
NOTE: A SSER value >= .80 is desired
************************************************.
```

Appendix B: SSER program for the one-way ANOVA

```
DATA LIST LIST / FOBS (F9.3) DF1 DF2 DF3 N (4F8.0).
*****************************************************************************
Between BEGIN DATA and END DATA below, put your observed F value (FOBS), the degrees of
freedom (DF1 for the numerator, DF2 for the denominator, and DF3 for the R^2 effect size where it is
always found through N-1) from your F test, and the sample size (N)
*****************************************************************************.
BEGIN DATA
4.623 1 38 39 40
END DATA.
*****************************************************************************
NOTE: Below in FDIFF for the critical value of F, choose the alpha level for a two-tailed test, either
FCRIT.05 or FCRIT.01. Currently, the program default is set at the .05 level
*****************************************************************************.
COMPUTE FCRIT.05 = ABS(IDF.F(.95,DF1,DF2)).
COMPUTE FCRIT.01 = ABS(IDF.F(.99,DF1,DF2)).
COMPUTE FDIFF = FOBS-FCRIT.05.
COMPUTE FREP = CDF.F(FDIFF,DF1,DF2).
COMPUTE R2 = FOBS/(FOBS+DF3).
COMPUTE FSIG = SIG.F(FOBS,DF1,DF2).
EXECUTE.
FORMAT FCRIT.05 TO FSIG  (F9.3).
VARIABLE LABELS R2 'R2 Effect Size (.10, .25, .40 are Suggested = Small, Medium, and Large
Effects)'/FOBS 'Your Observed F Value'/FCRIT.05 'For Your DF1 and DF2, the Critical Value of F,
Alpha=.05 (Program Default Value)'/FCRIT.01 'For Your DF1 and DF2, the Critical Value of F,
Alpha=.01'/DF1 'Degrees of Freedom for the Numerator'/DF2 'Degrees of Freedom for the
Denominator'/FSIG 'The Probability of the Observed F Value'/FREP 'The Upper Limit of the SSER
Probability Value'/.
REPORT FORMAT=LIST AUTOMATIC ALIGN (LEFT)
MARGINS (*,150)
  /VARIABLES= DF1 DF2 FOBS FCRIT.05 R2 FSIG FCRIT.01
  /TITLE "Test Statistics".
REPORT FORMAT=LIST AUTOMATIC ALIGN (LEFT)
MARGINS (*,110)
  /VARIABLES= FREP
  /TITLE "How Much Beyond a 50/50 Chance Do You Have of Replicating Your Statistically Significant
Findings?".
```

Appendix C: SSER program for chi-square

```
DATA LIST LIST / CHIOBS (F9.3) DF N (2F8.0).
***************************************************************************
Between BEGIN DATA and END DATA below, put your observed chi-square value (CHIOBS), the
degrees of freedom (DF), and the sample size (N)
***************************************************************************.
BEGIN DATA
4.32 1 40
END DATA.
***************************************************************************
NOTE: Below in CHIDIFF for the critical value of chi-square, choose the alpha level for either
CHICRIT.05, CHICRIT.01, or CHICRIT.001 Currently, the program default is set at the .05 level
***************************************************************************.
COMPUTE CHICRIT.05 = ABS(IDF.CHISQ(.95,DF)).
COMPUTE CHICRIT.01 = ABS(IDF.CHISQ(.99,DF)).
COMPUTE CHICRIT.001 = ABS(IDF.CHISQ(.995,DF)).
COMPUTE CHIDIFF = CHIOBS-CHICRIT.05.
COMPUTE CHIREP = CDF.CHISQ(CHIDIFF,DF).
COMPUTE C = SQRT(CHIOBS/(N+CHIOBS)).
COMPUTE SIG = 1-CDF.CHISQ(CHIOBS,DF).
EXECUTE.
FORMAT CHICRIT.05 TO SIG  (F9.3).
VARIABLE LABELS  C 'Pearsons Coefficient of Contingency (C) Effect Size (.10, .30, .50 are
Suggested = Small, Medium, and Large Effects)'/CHIOBS 'Your Observed Chi Square Value'/SIG 'The
Probability of Your Observed X2 Value'/CHICRIT.05 'For Your DF, the Critical Value of X2, Alpha=.05
(Program Default Value)'/CHICRIT.01 'For Your DF, the Critical Value of X2, Alpha=.01'/CHICRIT.001
'For Your DF, the Critical Value of X2, Alpha=.001'/DF 'Degrees of Freedom'/CHIREP 'The Upper Limit
of the SSER Probability Value'/.
REPORT FORMAT=LIST AUTOMATIC ALIGN (LEFT)
MARGINS (*,150)
  /VARIABLES= DF CHIOBS CHICRIT.05 SIG C CHICRIT.01 CHICRIT.001
  /TITLE "Test Statistics".
REPORT FORMAT=LIST AUTOMATIC ALIGN (LEFT)
MARGINS (*,110)
  /VARIABLES= CHIREP
  /TITLE "How Much Beyond a 50/50 Chance Do You Have of Replicating Your Statistically Significant
Findings?".
```