5-1-2009

Effects of Population Distribution, Sample Size and Correlation Structure on Huberty's Effect Size *R*

James B. Hittner
College of Charleston, hittnerj@cofc.edu

Effects of Population Distribution, Sample Size and Correlation Structure on Huberty's Effect Size R

James B. Hittner
College of Charleston

Huberty's (1994) R^2 is derived by subtracting the expected value of R^2 from an adjusted R^2, and the square root of Huberty's R^2 is Huberty's effect size R. The present study examined the effects of population distribution, sample size and population correlation structure on the statistical power of Huberty's R.

Key words: Huberty's R; Statistical Power; Multiple Regression; Effect Size.

Introduction

In the context of multiple regression analysis, it is often standard practice to examine whether the squared multiple correlation coefficient, R^2, is statistically significant. The intent of such a test is to determine whether R^2 differs significantly from zero, and the null hypothesis may be stated as H_o: $\rho^2 = 0$. Although this test is widely used, it is misleading because the expected value of R^2 is not zero when $\rho = 0$. Rather, as Morrison (1990) pointed out, the expected value, or expected long-run mean, of R^2 is equal to $p / N - 1$, where p is the number of predictor variables. The implication of this equation is that R^2 should be examined in relation to the expected value of R^2, $E(R^2)$, because the latter quantity is the value of R^2 that can be expected simply by chance.

In light of this realization, it seems more appropriate for researchers to test the null hypothesis, H_o: $\rho^2 = \rho_o^2$, where $\rho_o^2 = E(R^2)$. Darlington (1990) gave an F statistic for testing this null hypothesis and Huberty (1994) presented an adjusted R^2 index that takes into account the value of $E(R^2)$. The formula for

James Hittner is a Professor of psychology. His quantitative research interests include linear regression and correlation, confirmatory factor analysis, and statistical software development. Email: hittnerj@cofc.edu.

Huberty's adjusted R^2 index is:

$$R^2_{adj} = (R^2 - E(R^2)) / (1 - E(R^2)).$$

Huberty (1994) also presented an effect size measure for multiple regression studies that is calculated by subtracting $E(R^2)$ from Huberty's adjusted R^2 index. This effect size measure seems more appropriate than either R^2 or the adjusted R^2 given that it simultaneously accounts for both shrinkage and the sample size-to-predictor ratio.

Despite the apparent appropriateness of Huberty's effect size measure, standard statistical software packages, such as SPSS and Minitab, report only R^2 and adjusted R^2 values. Furthermore, although fourteen years have passed since the article was first published, very little, if any, quantitative research has been conducted on Huberty's proposed effect size measure. Due to this omission from the statistical literature, the present study generated simulated data and examined Huberty's effect size measure under different population distributions, sample sizes, and population correlation structures.

Methodology

Random variables were generated from the following three population distributions: Normal ($\mu = 0$, $\sigma = 1$), Weibull ($\lambda = 0.5$, $k = 1.2$), and Poisson ($\mu = \lambda = 0.5$). These distributions differ

in shape and are representative of the types of data distributions often encountered in applied research. The Weibull distribution, for example, is commonly used to model failure characteristics such as infant mortality, random failures, product wear-out, and the breaking strength of materials; it is also appropriate for lifetime modeling/survival analyses. Although similar in form to the exponential distribution the Weibull distribution can accommodate hazard changes over time, unlike the exponential which assumes a constant hazard rate (Heo, Faith, & Allison, 1998). The Poisson distribution is a discrete distribution that is often used to model counts, such as the number of arrivals, deaths, or failures in a given time period, and it can also be used to model the number of times a random event occurs over a given distance or across a particular spatial area. Such modeling of frequency count data per unit time, distance or area is tantamount to modeling rate data.

For each of the three population distributions, four random variables were generated for three different sample sizes (N's of 50, 100, 200) and three different population correlation structures (ρ's of 0.15, 0.30 and 0.65, representing low, moderate and high levels of correlation, respectively). This data generation process resulted in a total of 27 sets of four random variables (i.e., 3 distributions x 3 sample sizes x 3 correlation structures). For each set of four random variables, the specified correlation structure was induced by adding a multiple of a random variable, U, from the same population distribution to each randomly generated variable ($X1$, $X2$, $X3$, and Y). For each variable set, the value of the multiplicative constant, c, was chosen to produce the desired correlation. The specific algorithm was as follows:

$$X1\text{new} = (X1 + cU) / (1 + c^2)$$
$$X2\text{new} = (X2 + cU) / (1 + c^2)$$
$$X3\text{new} = (X3 + cU) / (1 + c^2)$$
$$Y\text{new} = (Y + cU) / (1 + c^2)$$

In generating the new, correlated, variables the choice as to which variable constituted Y was arbitrary. For consistency, the fourth correlated variable was always designated as Y. An important point concerning this

methodology is that the algorithm produces variables that correlate, on average, at the specified level of correlation. By generating variables that demonstrate approximate rather than exact and unvarying levels of correlation, the above algorithm produces sets of correlated variables that more closely mirror real-world datasets. For example, in the case of the Weibull distribution with $N = 200$ and a population correlation structure of 0.30, the mean empirical correlation for the four variables was 0.303 and the 95% confidence interval for the mean r ranged from 0.272 to 0.334. All of the variables in the present study were generated using the Statistical Package for the Social Sciences (SPSS, version 14).

For each of the 27 simulated datasets, a simultaneous multiple regression analysis was conducted (using SPSS) whereby Y was regressed onto the three predictor variables ($X1$, $X2$ and $X3$). The resulting R^2 value, along with the sample size, N, and the number of predictors, p, was then entered into a SAS data step program to calculate the expected value of R^2, Huberty's adjusted R^2 index, and Huberty's effect size measure (the SAS data step program is available from the author upon request). For each of the 27 datasets, the square root of Huberty's R^2 effect size measure - hereafter referred to as Huberty's effect size R - was examined to determine whether, given a specified sample size (50, 100, 200), number of predictors (3), level of statistical power (0.80) and alpha level (0.05), the value of R would be large enough to attain statistical significance at $p \leq 0.05$. The relevant power calculations were carried out using a FORTRAN program written by Dunlap, Xin, and Myers (2004). This program calculates power using the random, or unconditional, approach recommended by Gatsonis and Sampson (1989). Monte Carlo simulation results reported by Dunlap, et al. (2004) indicate that the random approach is more accurate than the more commonly used fixed approach. For each generated dataset, Huberty's effect size R was evaluated against the minimally detectable population R given the specified sample size, power = 0.80, alpha = 0.05, and $p = 3$ predictors. Based on Dunlap et al.'s power program, the minimally detectable population R values under these conditions for

N's of 200, 100, and 50 are 0.231, 0.323, and 0.448, respectively. Considering the above *R* values as comparative benchmarks, the objective of this study was to examine the effects of population distribution, sample size, and population correlation structure on the power of Huberty's effect size *R*, where power is defined as being adequate ($\geq$ 0.80) when Huberty's *R* exceeds the minimally detectable population *R*.

Results

For all cases with a correlation structure of ρ = 0.65, Huberty's effect size *R* estimates exceeded the minimally detectable population *R*, thereby demonstrating adequate levels of statistical power. For cases with a correlation structure of ρ = 0.30, six of the nine Huberty *R* estimates demonstrated adequate power, two demonstrated inadequate power, and one could not be calculated. The two underpowered cases were the Weibull distribution at *N* = 100 and the Poisson distribution at *N* = 100. The incalculable estimate was for the Weibull distribution at *N* = 50. Huberty's effect size *R* could not be computed for this case because the value of Huberty's adjusted R^2 index (0.030) was less than the expected value of R^2 (0.061). The difference between these two values equals Huberty's effect size measure, R^2, which in this case amounted to -0.031 (i.e., 0.030 - 0.061). Because the square root of a negative number cannot be computed, the value of Huberty's effect size *R* for this case is incalculable. For cases with a correlation structure of ρ = 0.15, six of the nine Huberty *R* estimates were underpowered and the remaining three could not be calculated (for the same reasons as noted above). The Huberty effect size *R* estimates and other relevant data for each case examined in this study are presented in Table 1.

One finding of interest concerns the two underpowered cases with a correlation structure of ρ = 0.30 (the Weibull and Poisson distributions at *N* = 100). In an effort to explain these findings, the mean empirical correlations and the coefficients of variation (CV; standard deviation of the empirical correlations divided by the mean correlation) for the Weibull and Poisson cases were compared against the corresponding, adequately powered, Normal

distribution case. All of the mean correlation comparisons were statistically nonsignificant (all Fisher *Z*-tests < 0.50, all *p*-values > 0.60). By contrast, all pairwise likelihood ratio tests on the CV's were statistically significant (*p*'s < 0.005), with the Weibull and Poisson CV's being significantly larger than the Normal distribution CV. These data suggest that, relative to the Normal case, the greater noise-to-signal ratio in the empirically generated correlations for the Weibull and Poisson cases may have contributed to their compromised levels of statistical power.

Another finding of interest was the negative value for Huberty's R^2 (and corresponding incalculable value for Huberty's effect size *R*) for the Weibull distribution at *N* = 50 and correlation structure of ρ = 0.30. Although the reason for this finding is not entirely clear, one possible explanation is that the small sample size (50) and relatively large CV (0.326) interacted with the shape (i.e., moments) of the Weibull distribution to produce an insufficiently large R^2 value. With respect to the cases with a correlation structure of ρ = 0.15, the fact that all six of the calculated Huberty *R* estimates were underpowered (three were incalculable) suggests that such a low level of intercorrelation among predictors and criterion generated a regression model that lacks adequate statistical power. It is important to note, however, that the data generation algorithm used in this study produced empirical correlations for the ρ = 0.15 cases that were noticeably more variable, as evidenced by the CV's, than were the correlations for the 0.30 and 0.65 cases. This heightened level of variability could have contributed to the underpowered estimates for the ρ = 0.15 cases. These same two factors (low level of intercorrelation, greater variability in estimated correlations), more so than sample size and distribution type, are the likely reasons underlying the incalculable Huberty *R* estimates.

One point worth mentioning about statistical power analysis in the context of multiple regression is that the algorithms used to compute integrals from the distribution of R^2 assume that the joint distribution of predictors and criterion is multivariate normal (Dunlap et al., 2004; Gatsonis & Sampson, 1989). When the multivariate distribution deviates from

Table 1: Huberty's Effect Size R and Related Statistics

Distrib	N	rstruct	ExpRsq	HuberRsq	HuberES	HuberESR	Meanr	Sdr	CVr
Weib	200	0.15	0.015	0.038	0.023	0.152	0.145	0.031	0.214
Weib	200	0.30	0.015	0.164	0.149	0.386	0.303	0.030	0.099
Weib	200	0.65	0.015	0.554	0.539	0.734	0.655	0.022	0.034
Poiss	200	0.15	0.015	0.033	0.018	0.134	0.139	0.048	0.345
Poiss	200	0.30	0.015	0.168	0.153	0.391	0.321	0.040	0.125
Poiss	200	0.65	0.015	0.557	0.542	0.736	0.665	0.022	0.033
Norm	200	0.15	0.015	0.023	0.008	0.089	0.156	0.064	0.410
Norm	200	0.30	0.015	0.140	0.125	0.354	0.321	0.058	0.181
Norm	200	0.65	0.015	0.519	0.504	0.710	0.659	0.034	0.052
Weib	100	0.15	0.030	0.001	-0.030		0.146	0.092	0.630
Weib	100	0.30	0.030	0.099	0.068	0.261	0.295	0.078	0.264
Weib	100	0.65	0.030	0.496	0.465	0.682	0.643	0.043	0.067
Poiss	100	0.15	0.030	0.012	-0.018		0.153	0.063	0.412
Poiss	100	0.30	0.030	0.123	0.093	0.305	0.312	0.060	0.192
Poiss	100	0.65	0.030	0.508	0.478	0.691	0.651	0.034	0.052
Norm	100	0.15	0.030	0.056	0.026	0.161	0.145	0.061	0.421
Norm	100	0.30	0.030	0.244	0.214	0.463	0.357	0.051	0.143
Norm	100	0.65	0.030	0.594	0.563	0.750	0.672	0.030	0.045
Weib	50	0.15	0.061	-0.054	-0.115		0.165	0.118	0.715
Weib	50	0.30	0.061	0.030	-0.031		0.304	0.099	0.326
Weib	50	0.65	0.061	0.491	0.430	0.656	0.668	0.043	0.064
Poiss	50	0.15	0.061	0.113	0.051	0.226	0.152	0.171	1.125
Poiss	50	0.30	0.061	0.315	0.254	0.504	0.390	0.128	0.328
Poiss	50	0.65	0.061	0.651	0.589	0.767	0.693	0.074	0.107
Norm	50	0.15	0.061	0.076	0.015	0.122	0.161	0.104	0.646
Norm	50	0.30	0.061	0.274	0.212	0.460	0.369	0.090	0.244
Norm	50	0.65	0.061	0.608	0.547	0.740	0.688	0.042	0.061

Notes: Distrib = Population distribution (Weibull, Poisson, Normal); N = Population sample size; rstruct = Population correlation structure; ExpRsq = Expected value of R^2; HuberRsq = Huberty's adjusted R^2; HuberES = Huberty's adjusted R^2 minus the expected value of R^2; HuberESR = The square root of HuberES; Meanr = Arithmetic average of empirically generated correlations (i.e., correlations among $X1$, $X2$, $X3$, and Y); Sdr = Standard deviation of empirically generated correlations; CVr = Coefficient of variation for empirically generated correlations (i.e., Sdr / Meanr). Blank entries for HuberESR indicate incalculable values (see text for details).

normality, then power estimates may become biased. However, the extent of bias is difficult to quantify and represents an important topic for future research. Another point worth noting is that the present investigation focused solely on factors affecting the power of Huberty's overall multiple regression coefficient. Factors affecting the power of individual predictors within the context of a larger regression model were not considered (for a treatment of this topic, the reader is referred to Maxwell, 2000). Though it is commonplace in multiple regression to test the partial contribution of a single predictor in the context of other predictor variables, such a practice is not without interpretive problems (Dunlap & Landis, 1998). A final point is that, in the present study, datasets with known *a priori* properties, in terms of population distribution, sample size and correlation structure, were generated and the obtained power of Huberty's effect size R was examined for each generated dataset. The present investigation was not a Monte Carlo simulation study in which the empirical properties of one or more statistical tests were examined. Such Monte Carlo work designed to investigate the power and efficiency (Type I error rate) of a significance test of Huberty's R represents an important direction for future research.

Conclusion

This study examined the power of Huberty's effect size R under three different population distributions (Weibull, Poisson, Normal), sample sizes (N's of 50, 100, 200), and population correlation structures (ρ's of 0.15, 0.30, and 0.65). For all conditions with a correlation structure of 0.65, Huberty's R demonstrated adequate statistical power. For cases with a correlation structure of 0.30, six of the eight estimated Huberty R values maintained adequate power (one value could not be calculated). For cases with a correlation structure of 0.15, the Huberty R values were either underpowered (six cases) or incalculable (three cases).

These results suggest that - in the context of multiple regression research - Huberty's effect size R maintains adequate statistical power under a variety of distributional shapes, samples sizes and correlation structures. The notable exception to this rule concerns cases with a correlation structure of 0.15, in which all of the estimated Huberty R values (six of nine cases) were underpowered. Such low power estimates suggest that practitioners of multiple regression analysis should restrict their attention to variables that correlate above 0.15 if they hope to maintain adequate statistical power for Huberty's effect size R (at least for models with 3 predictors and sample sizes $\leq$ 200). The precise *magnitude* of correlation needed to maintain adequate power for Huberty's R under various distributional shapes and sample size conditions is a topic for future research. It is hoped that the present study fosters a greater appreciation of Huberty's R and that the findings motivate additional research into factors that influence the statistical power of Huberty's effect size R.

References

Darlington, R. B. (1990). *Regression and linear models*. NY: McGraw-Hill.

Dunlap, W. P., & Landis, R. S. (1998). Interpretations of multiple regression borrowed from factor analysis and canonical correlation. *Journal of General Psychology*, *125*, 397-407.

Dunlap, W. P., Xin, X., & Myers, L. (2004). Computing aspects of power for multiple regression. *Behavior Research Methods, Instruments, & Computers*, *36*, 695-701.

Gatsonis, C., & Sampson, A. R. (1989). Multiple correlation: Exact power and sample size calculations. *Psychological Bulletin*, *106*, 516-524.

Heo, M., Faith, M. S., & Allison, D. B. (1998). Power and sample size for survival analysis under the Weibull distribution when the whole lifespan is of interest. *Mechanisms of Ageing and Development*, *102*, 45-53.

Huberty, C. J. (1994). A note on interpreting an R^2 value. *Journal of Educational and Behavioral Statistics*, *19*, 351-356.

Maxwell, S.E. (2000). Sample size and multiple regression analysis. *Psychological Methods*, *5*, 434-458.

Morrison, D. F. (1990). *Multivariate statistical methods*. NY: McGraw-Hill.