**Effect of *n1*/*n2* on Estimated *d* and *rpb*: Case in Point**

 Jenna Hartinger found a significant correlation between scores on the WART (a continuous measure of workaholism) and self-report of family history of high cholesterol (dichotomous, no or yes). She reported the point-biserial *r*, which was .22. By the usual conventions for interpreting *r*, this would be considered to be a small (.1) to medium (.3) sized correlation. One should, however, consider the fact that the value of the point biserial *r* is greatly influenced by the ratio of the sample sizes. When there are a lot more scores at the one level of the dichotomous variable than at the other level, the value of *r* is depressed. Estimated Cohen’s *d* (the standardized difference in means) might be a better way to express the magnitude of the relationship between a dichotomous variable and a continuous variable. For Jenna’s data, notice that there are nearly 2.5 times more respondents reporting no family history of high cholesterol than there are reporting such a history.

| **Group Statistics** |
| --- |
|  | FAMCHOL | N | Mean | Std. Deviation | Std. Error Mean |
| Work Addiction Risk  | dimension1 | No | 140 | 2.5653 | .38354 | .03242 |
| Yes | 56 | 2.7492 | .32811 | .04385 |

| **Independent Samples Test** |
| --- |
|  | t-test for Equality of Means |
| t | df | Sig. (2-tailed) |
| Work Addiction Risk  | Equal variances assumed | -3.154 | 194 | .002 |
| Equal variances not assumed | -3.372 | 117.645 | .001 |

 From these statistics one can estimate Cohen’s *d* as being .4987 (.5 is the benchmark for “medium”), with a confidence interval extending from .18 to .81.

[Return to my discussion of *rpb* versus *d*](http://core.ecu.edu/psyc/wuenschk/StatHelp/d-r.htm)