**HOW TO CALCULATE EFFECT SIZES**

by Simon Moss

|  |
| --- |
| **Introduction** |

 Suppose a researcher wants to ascertain whether Echinacea—a product that is often consumed to treat colds—can improve mood. One thousand participants consume Echinacea for a month. Another thousand participants consume a placebo for a month. Then, after a month, participants are asked to indicate the extent to which they feel satisfied with life.



 Now suppose that a statistical technique—called an independent t-test—reveals that Echinacea, relative to the placebo, does improve satisfaction with life, t(278) = 34.01, p < 0.001. The p value is particularly low, indicating that Echinacea almost definitely improves satisfaction with life. Nevertheless, this very low p value does not indicate that Echinacea greatly improves satisfaction with life. The reason is that p values depend on two considerations.

* First, if the effect size—that is, the extent to which Echinacea improves satisfaction in life—is high, the p value will tend to be very low
* Second, if the sample size or number of participants is very high, the p value will also tend to be very low

 Therefore, this very low p value could be ascribed to a large effect size, a large sample size, or both. We thus cannot be certain the effect size is large. We cannot be certain that Echinacea greatly improves satisfaction in life. The very low p value merely indicates we can be quite sure that Echinacea does improves satisfaction in life: significant differences do not necessarily indicate consequential effects.

|  |
| --- |
| **Measures of effect size: A simple example** |

 In short, to gauge the extent to which Echinacea improves satisfaction in life, we cannot depend on t values, p values, or other tests of significance: these tests also depend on sample size. Instead, we need a measure that does not depend on sample size. Fortunately, researchers have developed a series of formulas that measure effect size—such as the degree to which which Echinacea improves satisfaction in life. These measures are independent of sample size.

 To illustrate, when researchers want to compare two conditions—such as participants who consumed Echinacea and participants who did not consume echinacea—on some numerical measure, they tend to conduct an independent t-test. In these circumstances, to compute the effect size, they would utilize the following formula:



 Specifically

* d is the symbol for effect size in these circumstances
* the numerator symbolizes the difference between the means of each condition
* the demoninator is roughly the average standard deviation in each condition

 Thus, the effect size is merely the difference between the life satisfaction of participants who consumed Echinacea and the life satisfaction of participants who consumed the placebo over the standard deviation within each condition. A d value of 1, therefore, would indicate the difference between the two means equals the standard deviation within groups. A d value of 0.5 would indicate the difference between the two means equals half the standard deviation within groups.

|  |
| --- |
| **Interpretation of effect size** |

 Now suppose you show the effect size is 0.4. How can you interpret this number? What do you conclude from this number?

**Determine whether the effect is small, medium, or large**

 To interpret effect size, you first need to ascertain whether this number is small, medium, or large. When you calculate d, values around 0.2 are deemed as small, values around 0.5 are deemed as medium, and values around 0.8 are demeed as large. In this instance, you would conclude the effect size is medium or perhaps slightly less.

 But, now you have merely converted an arbitary number, 0.4, to an arbitrary label, medium? What does this label imply. What does the statement “the effect size of Echinacea on satisfaction with life is medium” actually indicate? To answer this question, we need to understand the actual meaning of small, medium, and large.

**The meaning of small, medium, and large effects**

 To understand the meaning of small, medium, and large, we need standard examples of what small, medium, and large effects are. The following table summarizes some examples, benchmarks, or descriptions that illustrate small, medium, and large effects. To illustrate

* As the first column shows, the difference in height between 15 and 16 year old girls is equivalent to a small effect.
* Indeed, if 15 year old girls were standing next to one wall and 16 year old girls were standing next to another wall, the difference in height would not be conspicuous.
* In addition, about 58% of the 16 year old girls would be taller than would the average 15 year old girl
* Now skim the second and third column of this table
* After skimming these columns, you might now have developed an intuition around what a small, medium, and large effect represents.
* Roughly, a small effect is usually inconspicuous, a medium effect is obvious to an expert, and a large effect is obvious to anyone

|  |  |  |
| --- | --- | --- |
| Small effect | Medium effect  | Large effect |
| Difference in height between 15 and 16 year old girls  | Difference in height between 14 and 18 year old girls | Difference in height between 13 and 18 year old girls |
| The difference or effect is almost inconspicuous | The difference or effect is probably obvious to an expert  | The difference or effect is probably obvious to anyone |
| 58% of the participants in one condition—such as individuals who consumed Echinacea—exceed the mean of the participants in the other condition on, for example, life satisfaction  | 69% of the participants in one condition exceed the mean of the participants in the other condition | 79% of the participants in one condition exceed the mean of the participants in the other condition |

|  |
| --- |
| **Other measures of effect size** |

 Thus far, we have discussed only one measure of effect size: d. Whenever researchers plan to compare two conditions on one measure—and thus are likely to conduct an independent t-test—this measure is useful. Otherwise, researchers need to utilize other measures. The following table summarizes these measures of effect size (see Cohen, 1992). In particular, this table indicates when these measures are applicable and which levels represent small, medium, and large effects. Roughly, most of these measures assess the degree to which the variability in one variable vanishees after controlling another variable.

|  |  |  |
| --- | --- | --- |
| Measure of effect size  | Situations in which the measure is applicable  | Small, medium, and large  |
| d | Comparing two conditions on one measure | 0.2, 0.5, 0.8 |
| partial eta squared or η2.  | Comparing more than two conditions on one measure—for ANOVAs or ANCOVAs | 0.01, 0.06, 0.14 |
| correlation or r;semi-partial correlation  | exploring the association between two numerical measures | .1, .3, .5 |
| R2 | Multiple regression or SEM | .1, .9, .25 |
| f2 =R2 / (1- R2) | Multiple regression or SEM | .02, .15, .35 |
| odds ratio | Comparing two conditions on categorical measures  | 1.68, 3.47, 6.71 \* |

\* See Chen, Cohen, and Chen (2009).

 In more complex designs, other measures of effect size might need to be calculated:

* For instances in which the design comprises several measures, such as when MANOVAs are used, see Olejnik and Algina (2000)
* For repeated measures ANOVAs, see Bakeman (2005)
* For mult-level modelling, see a discussion by Tymms, Merrell, & Henderson (1997)
* For some other measures, see Snyder and Lawson (1993).

|  |
| --- |
| **How to calculate these measures** |

 The following table clarifies how to calculate these measures of effect size.

|  |  |
| --- | --- |
| Measure of effect size  | Situations in which the measure is applicable  |
| d | Some statistics programs might automatically generate d. Otherwise, you can easily calculate d. However, to calculate d most accurately* If the standard deviation differs between the two groups, calculate a statistic called the pooled standard deviation
* In particular, multiply each standard deviation by n – 1, where n is the sample size of that group.
* Divide the sum of these two values by the total number of participants – 2
* Square root this answer
* Below is the formula

After you have constructed this pooled standard deviation* Calculate the difference between your two means
* Divide by the pooled standard deviation
 |
| d for paired-samples t-tests | If the conditions are matched—or comprise the same participants—you need to utilize a different formula to compute d (Ray & Shadish, 1996). In particular* Locate the t value you generate
* Square root the number of individuals in each condition
* d = the first value / the second value
 |
| partial eta squared or η2 | Most statistics programs will calculate this partial eta squared. Or, to compute this effect* In an ANOVA table, locate the SS for the relevant main effect or interaction
* Locate the SS for the error term, sometimes called residual
* The formula is simply SS effect/(SS effect + SS error)
 |
| correlation or r | All statistics programs will calculate a correlation of course.  |
| R2 | All statistics programs will calculate this value |
| f2 =R2 / (1- R2) | Simply derive from R2 |

|  |
| --- |
| **Complications and clarifications** |

Here is some interesting insights about measures of effect size

* To calculate d, if researchers utilize the pooled standard deviation, as recommended earlier, the value they generate slightly overestimates the true effect size. Some books recommend a correction that offsets this bias (Hedges & Olkin, 1985, page 80)
* The effect size measures are not as meaningful when data are not normally distributed (see <https://www.leeds.ac.uk/educol/documents/00002182.htm>)

|  |
| --- |
| **References** |

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. Behavioral Research Methods, 37(3), 379-384.

Chen, H., Cohen, P., Chen, S (2009). How big is a big odds ratio? Interpreting magnitude of odds ratios in epidemiological studies Communication in Statistics—Simulation and Computation, 39(4)

Cohen, J. (1992). A power primer. Psychological Bulletin, 112(1), 155.

Hedges, L. and Olkin, I. (1985) Statistical Methods for Meta-Analysis. New York: Academic Press.

Olejnik, S., & Algina, J. (2000). Measures of effect size for comparative studies: applications, interpretations and limitations. Contemporary Educational Psychology, 25, 241-286.

Ray, J. W., & Shadish, W. R. (1996). How interchangeable are different estimators of effect size? Journal of Consulting and Clinical Psychology, 64, 1316-1325.

Snyder, P., & Lawson, S. (1993). Evaluating results using corrected and uncorrected effect size estimates. Journal of Experimental Education, 61, 4, 334-349.

Tymms, P., Merrell, C. and Henderson, B. (1997). The first year as school: a quantitative investigation of the attainment and progress of pupils. Educational Research and Evaluation, 3, 101-118.