The possible advantages of the mean absolute deviation ‘effect’ size

Stephen Gorard
Durham University
s.a.c.gorard@durham.ac.uk

A range of ‘effect’ sizes already exists, for presenting a relatively easy to interpret estimate of a difference or change between two sets of observations.

All are based on use of the standard deviation of the observations, involving squaring and then square-rooting, which makes results hard to interpret, hard to teach and may distort extreme scores.

An effect size based on the simpler mean absolute deviation overcomes these issues to some extent, while being at least as efficient and leading to the same substantive results in almost all cases.

This paper proposes the use of an easy to comprehend effect size based on the mean difference between treatment groups, divided by their mean absolute deviation. Using a simulation of 1,656 trials each of 100 cases using a before and after design, the paper shows that the substantive findings from any such trial would be the same whether a traditional effect size like Cohen’s d or the mean absolute deviation effect size is used. The mean absolute deviation effect size works. Among the advantages of using the mean absolute deviation effect size are its relative simplicity, efficiency, everyday meaning, and the lack of distortion of extreme scores caused by the squaring involved in computing the standard deviation. Given that working with absolute values is no longer the barrier to computation that it apparently was before the advent of digital calculators, there is a clear place for the mean absolute deviation effect size.

A range of existing effect sizes

In social science, as in natural and health sciences, the reporting of ‘effect’ sizes for numeric experimental and other empirical results is becoming more frequent. It is for example, the approach insisted on in the current publication manual of the American Psychological Association (2009) instead of significance testing. Significance testing is easily misunderstood, gives misleading results about the substantive nature of results, and is ‘best avoided’ (Lipsey et al. 2012, p.3). Effect sizes are often needed in situations involving population data or non-random samples where p-values based on probabilistic uncertainty would be entirely inappropriate anyway. In fact, the whole panoply of significance testing does not work as intended in real-life, and should cease (Gorard 2010). Instead a greater emphasis on straightforward reporting of results is needed, within a clear research design, and placed in context so that the scale of the results and the size and quality of the dataset can be judged.

There is a range of effect sizes for different types and distributions of data, including common variation, differences in variation, multiple groups, categorical variables and so on (Gorard 2013). This paper focuses on simple comparisons of means between two groups, assuming that the scores for both groups are for the same variable using the same scale of measurement. Conversion of the results into a standardised effect size might be done in order to help readers understand the substantive importance of the result or to allow the result to be synthesised with results from other studies using different measures. A common method of creating a standard effect size is to divide the difference between two means by their standard deviation (SD). In theory the SD to use here is that for the whole population. If only sample figures are available then the sample SD can be used instead, but even this compromise is ambiguous and the subject of much dispute. If the experiment has a pre-test then the SD of the pre-test scores for both groups uses the largest number of cases that are unaffected by the experimental intervention. There is an inevitable delay with maturation between the pre-test and the post-test, and the pre-test is rarely exactly the same as the post-test to prevent practice effects. These mean that the SD of the sample at pre-test may be a poor estimate of the SD of the population at post-test. Another possibility is to use the SD of the control group post-test scores. These are similarly unaffected by the intervention, presumably, and are more relevant to consideration of the outcome effect size. Unfortunately, the number of cases will inevitably be smaller than the combined total. So perhaps the best estimate of the SD will come from either the SD of the overall post-test scores, or the pooled SDs of the treatment and control group post-test scores. Given these and other variations, the standard effect size of difference between means divided by their standard deviation is not really that standard, with Cohens' d, Glass's delta, and Hedges' g and others all giving similar but slightly different final results from the same datasets. This paper suggests a new and similar variation on the effect size, to add to these, which has the advantage of being easier to comprehend for a wide audience.

The mean absolute deviation effect size

A further alternative would be to use the mean absolute deviation as the measure of dispersion to create a mean absolute deviation effect size. For a simple experimental design, this would be the difference between the mean outcome scores for both groups divided by the mean absolute deviation of the scores (for the pre-test, control group, or pooled groups etc.). The illustration that follows is based on a simulation, involving 1,656 pairs of sets of random numbers between 0 and 1. Each pair is envisaged as being the before and after scores for a set of 100 cases. Each pair yields a gain score, and a mean absolute deviation and standard deviation for each gain score column. These are correlated with each other using Pearson's R.

The mean absolute deviation (M|D| or perhaps just 'a' for simplicity) itself is an increasingly relevant alternative to the more common standard deviation (SD) as a measure of dispersion (Gorard 2005). M|D| is simply the average of the absolute differences between each score and the overall mean. When working with a true random sample and a population, the SD is calculated in the same way for both groups, and so is M|D|. Both are good summaries of the sample information. They are equally consistent and sufficient (Fisher 1920). It has now been shown that M|D| is at least as efficient as SD, and usually has the smallest probable error as an estimate of the equivalent population parameter (Stigler 1973). The standard deviation is only better under ideal conditions, working with a perfect normal distribution and no errors or missing data (Tukey 1960). And all of this is irrelevant anyway for most social scientists working with population data or convenience samples. In real-life research M|D| is to be preferred (Barnett and Lewis 1978, p.159, Huber 1981, p.3). It is preferred because it is more efficient in practice, gives each deviation its proportionate place in the result, and is easier for new researchers and others to understand - largely because it does not require the squaring and square rooting of differences. This is clear from experience of teaching, the unpopularity of the topic among social science students (Murtonen and Lehtinen 2003), and consideration of the wider readership for social science. For example, almost everyone who can see how to split a restaurant bill into equal shares (the mean) for seven friends can also judge whether this is unfair because one or more of these friends incurred costs that deviate markedly from that mean. They will, almost unconsciously, calculate the mean (absolute) deviation and use it to make a real-life decision about whether splitting the bill is fair. Asking them to square the deviations before summing them and then square rooting the result is much more complex, and sounds and is in the context quite ridiculous.

All of this has led to the mean absolute deviation being used routinely in a number of areas, including astronomy biology, engineering, IT, physics, imaging, geography and environmental science (e.g. Eddington 1914 p.147, Anand and Narasimha 2013, Hao et al. 2012, Hižak J. and Logožar R. 2011, Sari et al. 2012). For most authors the use of absolute numbers is no longer a barrier to computation. It would, of course, help if major analytical software such as SPSS included a routine to calculate M|D| (but SPSS still does not even have a routine for calculating the kind of effect sizes discussed in this paper). M|D| is linked to a range of other simple analytical techniques, again with relatively easy to understand meanings, including a 'segregation index' (GS) for summarising the unevenness in the distribution of individuals between organisational units (Gorard and Taylor 2002), and the relative difference, or achievement gap (Gorard et al. 2001). The purpose of this paper is not to revisit these claims and findings but to propose an effect size based on M|D|.

The important choice of whether the M|D| or SD is most suitable as a measure of dispersion then relates to which should be used in calculating an 'effect' size. Table 1 shows the relationship between the two versions of the effect sizes for 1,656 simulated trials, using both standard deviation and mean absolute deviation. To three decimal places, the SD effect size is indistinguishable from the M|D| effect size. With data such as these, it really does not matter which is used, and they will yield the same substantive conclusion in practice. This shows that the M|D| effect size is at least as useful as any of the alternatives above. It gives a 'sensible' result, but will be more tolerant of outliers, and is easier to comprehend.

Table 1 – Correlation between five experimental outcomes

	Standard deviation	Mean |deviation|	SD 'Effect' size	M|D| 'Effect' size
Standard deviation		+.953	-.005	-.006
Mean |deviation|	+.953		-.013	-.014
SD 'Effect' size	-.005	-.013		1
M|D| 'Effect' size	-.006	-.014	1	

N=1,656

Conclusion

As a measure of dispersion and as the denominator for calculating effect sizes, the standard deviation has a big advantage. It is already in widespread use – by definition it is linked to the normal distribution which also means that it appears in many statistical settings and guises. This is an important factor when selecting a standard effect size. However, it is hard to teach to new researchers, and has no easy to understand meaning in real-life. It also exaggerates the importance of extreme scores for no clear reason, so promoting the routine deletion of purported ‘outliers’, and is less efficient than the mean absolute deviation in the realistic situation where data is not in an ideal normal distribution, or where it has any errors at all.

Therefore, the mean absolute deviation effect size is proposed here. It is easier than the SD effect size for everyone to understand with an almost everyday meaning. Like the SD, the mean absolute deviation is already in use in a variety of fields and that use is growing. A new form of statistics with an absolute mean deviation effect size, absolute mean deviation correlation, least deviation regression models and so on is possible. In many contexts this new kind of statistics would be more robust (Amir 2012), and could fit together better (Cahan and Gamliel 2011). The computational problems with absolute numbers have been effectively solved by the power of modern computing. The main remaining drawback to the use of a metric with absolute numbers in it therefore concerns algebraic manipulation. However, most potential analysts do not want to carry out any algebraic manipulation. Most might prefer descriptive statistics, whether for populations, samples or simply groups of cases, which are more democratic than they are currently.

References

American Psychological Association (2009) *Publication manual of the APA 6th Edition*, Washington DC: American Psychological Association

Amir, E. (2012) On uses of mean absolute deviation: decomposition, skewness and correlation coefficients, *METRON – International Journal of Statistics*, LXX, 2-3, 145-164

Anand, M. and Narasimha, Y. (2013) Removal of salt and pepper Noise from highly corrupted images using mean deviation statistical parameter, *International Journal on Computer Science and Engineering*, 5, 2, 113-119

Barnett, V. and Lewis, T. (1978) *Outliers in statistical data*, Chichester: John Wiley and Sons

Cahan, S. and Gamliel, E. (2011) First among others? Cohen’s d vs. alternative standardized mean group difference measures, *Practical Assessment, Research and Evaluation*, 16, 10, 1-6

Eddington, A. (1914) *Stellar movements and the structure of the universe*, London: Macmillan

Fisher, R. (1920) A mathematical examination of the methods of determining the accuracy of observation by the mean error and the mean square error, *Monthly Notes of the Royal Astronomical Society*, 80, 758-770

Gorard, S. (2005) Revisiting a 90-year-old debate: the advantages of the mean deviation, *The British Journal of Educational Studies*, 53, 4, 417-430

Gorard, S. (2010) All evidence is equal: the flaw in statistical reasoning, *Oxford Review of Education*, 36, 1, 63-77

Gorard, S. (2013) *Research Design: Robust approaches for the social sciences*, London: SAGE

Gorard, S. and Taylor, C. (2002) What is segregation? A comparison of measures in terms of strong and weak compositional invariance, *Sociology*, 36, 4, 875-895

Gorard, S., Rees, G. and Salisbury, J. (2001) The differential attainment of boys and girls at school: investigating the patterns and their determinants, *British Educational Research Journal*, 27, 2, 125-139

Hao, Y., Flowers, H., Monti, M. and Qualters, J. (2012) U.S. census unit population exposures to ambient air pollutants, *International Journal of Health Geographics* 2012, 11, 3 doi:10.1186/1476-072X-11-3

Hižak J. , Logožar R. (2011) A derivation of the mean absolute distance in one-dimensional random walk, *Technical Journal*, 5, 1, 10-16

Huber, P. (1981) *Robust Statistics*, New York: John Wiley and Sons

Lipsey, M., Puzio, K., Yun, C., Hebert, M., Steinka-Fry, K., Cole, M., Roberts, M., Anthony, K. and Busick, M. (2012) *Translating the statistical representation of the effects of education interventions into more readily interpretable forms*, Washington DC: Institute of Education Sciences

Murtonen, M. and Lehtinen, E. (2003) Difficulties experienced by education and sociology students in quantitative methods courses, *Studies in Higher Education*, 28, 2, 171-185

Sari, S., Roslan, H. and Shimamura, T. (2012) *Noise estimation by utilizing mean deviation of smooth region in noisy image*, Fourth International Conference on Computational Intelligence, Modelling and Simulation, http://eprints.uthm.edu.my/3265/1/4871a232.pdf

Stigler, S. (1973) Studies in the history of probability and statistics XXXII: Laplace, Fisher and the discovery of the concept of sufficiency, *Biometrika*, 60, 3, 439-445

Tukey, J. (1960) A survey of sampling from contaminated distributions, in Olkin, I., Ghurye, S. , Hoeffding, W., Madow, W and Mann, H. (Eds.) *Contributions to probability and statistics: essays in honor of Harold Hotelling*, Stanford: Stanford University Press