International Refereed Journal of Engineering and Science (IRJES)
ISSN (Online) 2319-183X, (Print) 2319-1821
Volume 1, Issue 4(December 2012), PP.27-32
www.irjes.com

Computation and Interpretation of Effect Size in Significance Test

Habib Ahmed Elsayir

(PhD Statistics) Umm Al Qura University(K.S.A) Al Qunfudha University College
Dept.of Mathematics, Qunfudha

Abstract: *Effect size, a simple way of quantifying the difference between two groups, is a very important task in research problems and statistical significance tests . This paper discusses the issue of effect size, explains what an effect size is, how it is calculated and how it can be interpreted. Some related problems such as statistical significance, meta-analysis and power analysis are described. An implementation on theoretical data would be carried .Results , conclusions, recommendations on the use of effect sizes would be summarized.*
Key Words*: *Effect Size, Meta-analysis, , Sample size, Significance Test.

I Introduction

Effect size is an important tool in reporting and interpreting effectiveness ,and has many advantages over the use of tests of statistical significance. 'Effect size' is valuable for quantifying the effectiveness of a particular intervention, relative to some comparison, and a one of the tools that will help researchers move beyond null hypothesis testing. This paper gives a brief demonstration of basic methodologies of effect size, reviews issues of the topic, accompanied by numerical illustrations. Tables made are computed from different sources and verified using online software on effect size (see the list of websites references here) .The use of effect sizes, however, has generally been limited to meta-analysis - for combining and comparing estimates from different studies. This is despite the fact that measures of effect size have been available for decades , Huberty (2002). The concept of effect size is derived from a school of methodology known as meta-analysis , (see Baker,R.&Dwyer,F.(2000),Biostat(2006),Poston,J.M.,&Hanson,W.E.(2010).Heavily laying on Rosenthal (1994), Rosenthal &Rosnow(2000), has introduced a useful summary of effect sizes computation and transformations for inferential statistics. Michael Fur(2008) has also discussed effect sizes and their links to inferential statistics.

II Effect Size Computation formulas

The developed formulas for effect size calculation vary depending on whether the researcher plans to use analysis of variance(ANOVA), t test, regression or correlation, (see Morris and DeShon's (2002))..Formulas used to measure effect size can be computed in either a standardized difference between two means, or in the correlation between the independent variable classification and the individual scores on the dependent variable, which is called the "effect size correlation" (Rosnow & Rosenthal(1996).

Effect size for differences in means is given by Cohen's" d"Cohen, J. (1988) , is defined in terms of population means (μs) and standard deviation (σ), as shown below:

$$d = \frac{|\mu_1 - \mu_2|}{\sigma} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

There are several different ways that one could estimate σ from sample data which leads to multiple variants within the Cohen's d family.(see Karl L. Wuensch(2010)).
When using the root mean square standard deviation, the" d " is given as:

$$d = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{s^2_1 + s^2_2}{2}}} \ldots\ldots\ldots\ldots\ldots\ldots (2)$$

A version of Cohen's d uses the pooled standard deviation and is also known as Hedges' g(see is :

$$d = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{(n_1 - 1)S^2_1 + (n_2 - 1)S^2_2}{n_1 + n_2 - 2}}} \ldots\ldots (3)$$

The value can be obtained from an ANOVA program by taking the square root of the mean square error which is also known as the root mean square error.
Another model of Cohen's " d" using the standard deviation for the control group is also known as Glass' Δ (see Karl L. Wuensch(2010)),where:

$$d = \frac{|\bar{x}_1 - \bar{x}_2|}{s_{cont}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots . (4)$$

The control group's standard deviation is used because it is not affected by the treatment .It is suggested to use a pooled within group standard deviation because it has less sampling error than the control group standard deviation such that equal size constrain is adopted. When there are more than two groups , the difference between the largest and smallest means divided by the square root of the mean square error will be used ,i.e:

$$d = \frac{\bar{x}_{largest} - \bar{x}_{smallest}}{\sqrt{mse}} \quad \ldots\ldots\ldots\ldots\ldots .. (5)$$

As For OLS regression the measure of effects size is F which is defined by Cohen as follows:

$$f^2 = \frac{\rho^2}{1 - \rho^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (6)$$

Or ,as usually computed by taking the square root of f2
Once again there are several ways in which the effect size can be computed from sample data. It can be noted that η2 is another name for R2,the coefficient of determination, where: (see Karl L. Wuensch(2010)).

$$f^2 = \frac{R^2}{1-R^2} = \frac{eta^2}{1-eta^2} \ldots\ldots\ldots\ldots\ldots(7)$$

The effect size used in analysis of variance is defined by the ratio of population standard deviations:

$$f = \frac{\sigma_{means}}{\sigma} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (8)$$

Based on definitional formula in terms of population values., effect size w can be viewed as the square root of the standardized chi-square statistic.

$$w = \sqrt{\sum \frac{(\pi_0 - \pi_1)^2}{\pi_0}} \quad \ldots\ldots\ldots\ldots\ldots . (9)$$

And w is computed using sample data by the formula:

$$w = \sqrt{\sum \frac{(P_0 - P_1)^2}{P_0}} \quad \ldots\ldots\ldots\ldots\ldots . (10)$$

According to Poston &Hanson(2010),when a study reports a hit rate (percentage of success after taking the treatment or no treatment),the following formula can be used:

d= arcsine(p1)+ arcsine(p2)

Where p1 and p2 are the hit rates of the two groups.
If the effect size estimate from the sample is d, then it is normally distributed, with standard deviation:

$$\sigma(d) = \sqrt{\frac{N_{exp} + N_{cont}}{(N_{exp})\,(N_{cont})} + \frac{d^2}{2\,(N_{exp} + N_{cont})}} \quad \ldots\ldots\ldots(11)$$

(Where N exp and Ncont are the numbers in the experimental and control groups, respectively.)
The control group will provide the best estimate of standard deviation, since it consists of a representative group of the population who have not been affected by the experimental intervention. Therefore, it is often better to use a 'pooled' estimate of standard deviation ,which is given by

$$SD(pooled) = \sqrt{\frac{(N_{exp}-1)\ SD^2{}_{exp} + (N_{cont}-1)\ SD^2{}_{cont}}{(N_{exp}) + (N_{cont}) - 2}} \qquad \text{………(12)}$$

(Where Nexp and Ncont are the numbers in the experimental and control groups, respectively, with their variances.)

Based on the above formulas values, the larger the effect size, the greater is the impact of an intervention. Cohen suggested that a correlation of 0.5 is large, 0.3 is moderate, and 0.1 is small Cohen defined .40 as the medium effect size because it was close to the average observed effect size(Aguinis, & Harden(2009). The usual interpretation of this statement is that anything greater than 0.5 is large, 0.5-0.3 is moderate, 0.3-0.1 is small, and anything smaller than 0.1 is trivial.

2.1.Effect Size , Significance and Meta-analysis

Effect size is a name given to a set of indices that measure the magnitude of a treatment effect. Unlike significance tests, these indices are independent of sample size. Effect size measures are the common currency of meta-analysis studies that summarize the findings from a specific area of research. Effect size quantifies the size of the difference between two groups, and may therefore be said to be a true measure of the significance of the difference. Another use of effect size is its use in performing power analysis, (see Buchner,A., Erdfelder,E. and Faul,F(2009) . Researcher designers use power analysis to minimize the likelihood of both false positives and false negatives (Type I and Type II errors, respectively) ,Richard A. Zeller and Yan Yan (2007).

A number of statistics are sometimes proposed as alternative measures of effect size, other than the 'standardized mean difference'. One of these is the Proportion of variance accounted for, the R2 which represents the proportion of the variance in each that is 'accounted for' by the other. There are also effect size measures for multivariate outcomes. A detailed explanation can be found in Olejnik and Algina (2000). Calculating effect size is important when testing the goodness fit ,or contingency test,. For this test, the effect size symbol is w. Once effect size is known, this information can be used to calculate the number of participants needed and the critical chi-square value (for sample size rules (see Aguinis, H. & Harden, E. E. (2009)),(and see the effect of sample size on effect size in Slavin, R., & Smith, D. (2008).

III Effect Size Computation

The interpretations of effect-sizes given in Table (1) , in which a suggested values for low, medium and high effects is given, depend on the assumption that both control and experimental groups have a 'normal' distribution, other wise, it may be difficult to make a fair comparison between an effect-size based on normal distributions and one based on non-normal distributions. In practice, the values for large effects may be exceeded with values Cohen's d greater than 1.0 not uncommon.

Considering table(1) and table (2),it can be noted that, d can be converted to r and vice versa. For example, the d value of 0.8 corresponds to an r value of 0.371.The square of the r-value is the percentage of variance in the dependent variable that is accounted for by the effect in the explanatory variable groups. For a d value of 0.8, the amount of variance in the dependent variable by membership in the treatment and control groups is 13.8%.T-tests are used to evaluate the null hypothesis. For this test, the effect size symbol is r. If the desired effect size is known, statistical power and needed sample size can be calculated. For instance, if the target is to find how many elements are need in a study for a medium effect size (r = 0.30) with an alpha of .05. and power of 0.95, this information can be used to find the answer. For ANOVA, the effect size index f is used, and the effect size index from the group means can then be computed.

Power is the chance that if "d" exists in the real world, one gets a statistically significant difference in the data .if the power level is taken to be 80%,there is an 80% chance to discover a really existing difference in the sample. Alpha is the chance that one would conclude that an effect difference "d", has been discovered, while in fact this difference or effect does not exist. If alpha is set at 5%, this means that in 5%, or one in twenty, the data indicate that "something" exists, while in fact it does not.In table (3),consider that: power = 1-β = p (HA is accepted/HA is true).Set α ,the probability of false rejecting Ho,equal to some small value .Then ,considering the alternative hypothesis HA,choose a region of rejection such that the probability of observing a sample value in that region is less than or equal to α when Ho is true. If the value of sample statistic falls within the rejection region, the decision is made to reject the null hypothesis. Typically is set at 0.05, and critical t values are specified. The calculation works as follows: Entering α=0.05, power=0.95, effect size specified as in column (1), we find the needed elements (sample size (column 4)) and so on. The effect size is seen in table (3) Column(1). The effect size conventions are small =0.20, medium=0.50, large=0.80. Calculate d and r using t values and df (separate groups t test)calculate the value of Cohen's d and the effect size correlation r ,using the t test value for a between subjects t test and the degrees of freedom. Results are shown in table (4),while in table (5) ,d and r are calculated using t values and df.

IV Discussion

The effect size refers to the magnitude of the effect under the alternative hypothesis. It should represent the smallest difference that would be of significance. It varies from study to study. It is also variable from one statistical procedure to the other. It could be the difference in cure rates, or a standardized mean difference or a correlation coefficient. If the effect size is increased, the type II error decreases. Power is a function of an effect size and the sample size. For a given power, 'small effects' require larger sample size than 'large effects'. Power depends on (a) the effect size, (b) the sample size, and (c) the significance level. But if the researcher knew the size of the effect, there would be no reason to conduct the research.To estimate a sample size prior to doing the research, requires the postulation of an effect size, which might be related to a correlation, an f-value, or a non-parametric test. In the procedure implemented here ,'d' is the difference between two averages, or proportions. Effect size 'd' is mostly subjective, it is the difference you want to discover as a researcher or practitioner and it is a difference that you find relevant. However, if cost aspects are included ,'d' can be calculated objectively. The size of the difference in the response to be detected, which relates to underlying population, not to data from sample, is of importance since it measures the distance between the null hypothesis (HO) and specific value of the alternative hypothesis (HA).A desirable effect size is the degree of deviation from the null hypotheses that is considered large enough to attract the attention. The concept of small, medium, and large effect sizes can be a reasonable starting point if you do not have more precise information. (Note that an effect size should be stated in terms of a number in the actual units of the response, not a percent change such as 5% or 10 %.). Sample size determination and power analysis involve steps that are fundamentally the same .These include the investigation of; type of analysis and null hypothesis; power and required sample size for a reasonable range of effect as well as calculation of the sample size required to detect a reasonable effect with a reasonable level of power .Although effect size is a simple and readily interpreted measure of effectiveness, it can also be sensitive to a number of spurious influences, so some care needs to be taken in its use. Some of these issues are the form of the standard deviation used, and the normality assumption

V Conclusion

On the use of effect-sizes ,the following can be summarized:

i. Effect size is generally useful for quantifying effects for comparing the relative sizes of effects from different studies.
ii. Assumptions about the population nature is essential in using effect size , for the interpretation depends mainly on the assumptions of normality and equality of deviations of 'control' and 'experimental' group values. Effect sizes can be interpreted in terms of the percentiles or ranks at which two distributions overlap.
iii. Use of an effect size with a confidence interval holds the same information as a test of statistical significance, but with the emphasis on the significance of the effect, rather than the sample size.

References

[1.] Aguinis ,H.&Haren,E.E.(2009).Sample size rules of thumb: Evaluating three common practices. In Charles E. Lance and Robert J. Vandenberg (Eds.),Statistical and methodological myths and urban legends:Doctrine,verify and fable in the organizational and social sciences (pp.267-286).New York:Routledge.

[2.] Baker,R.,&Dwyer,F.(2000).A meta-analytic assessment of the effects of visualized instruction, paper presented at the 2000 AECT National Convention. Long Beach,CA.

[3.] Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates .

[4.] Huberty, C.J.. (2002) 'A history of effect size indices'. Educational and Psychological Measurement, 62, 2, 227-240.

[5.] Mike Fur(2008) Effect sizes and significance Tests .Psychology Dept. Wake Forest University(.March,2008).

[6.] Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. Psychological Methods, 7, 105-125.

[7.] Olejnik, S. and Algina, J. (2000) 'Measures of Effect Size for Comparative Studies: Applications, Interpretations and Limitations.' Contemporary Educational Psychology, 25, 241-286.

[8.] Poston,J.M,&Hanso,W.E.(2010).Meta-analysis of psychological assessment as a therapeutic intervention. Psychological Assessment ,22,203-12.

[9.] Richard A. Zeller and Yan Yan (2007). Establishing the Optimum Sample Size .Epidemiology and Medical Statistics ,Handbook of Statistics Volume 27, Pages 656-678.

[10.] Rosenthal, R. (1994) 'Parametric Measures of Effect Size' in H. Cooper and L.V. Hedges (Eds.), The Handbook of Research Synthesis. New York: Russell Sage Foundation.

[11.] Rosenthal, R.Rosnow,R.L.&Rubin (2000)D.B. Contrasts and Effect sizes in behavioral research :A correlational approach. Cambridge .UK. Cambridge University Press.

[12.] Slavin,R.,&Smith,D.(2008).Effects of sample size on effect size in systematic reviews in education .Paper presented at the annual meetings of the Society for Research on Effective Education, Crystal City ,VI

Websites :

-Becker,L.,(2000) . Effect Size (ES), This website provides an overview about effect size and its measures (including several effect size calculators).
-Biostat (2006). Meta-analysis. [On-line] Available: http://www.meta-analysis.com/.
-.Buchner,A., Erdfelder,E. and Faul,F. How to use G*power. Viewed 19 Jun 2009 http://www.psycho.uni
- duesseldorf.de/aap/projects/gpower/how_to_use_gpower.html.
-Karl L. Wuensch(2010), Standardized Effect Size Estimation: Why and How? Wuensch's Statistics Help Page. http://core.ecu.edu/psyc/ Wuenschk/stathelp/stathelp.htm.
-.Simon, Steve(1999)."Type I &Type II Error. Educational Statistics, Discussion List(EDSTAT-L(online). Available E-mail: edstat-l@jse.stat.ncsu.edu [1999, January 7 .

Tables :

Table(1): Effect size levels for different tests:

Large	medium	small		
0.80	0.50	0.20	d	t-test for means
0.50	0.30	0.10	r	t-test for correlation
0.35	0.15	0.02	f^2	F-test for regression
0.40	0.25	0.10	f	F-test for ANOVA
0.50	0.30	0.10	w	chi-square

Table(2):Effect sizes for two groups

f^2	f	r^2	r	d
0.999396	0.999698	0.49985	0.707	2
0.810155	0.900086	0.44756	0.669	1.8
0.641026	0.800641	0.39063	0.625	1.6
0.488824	0.699160	0.32833	0.573	1.4
0.359058	0.599214	0.26420	0.514	1.2
0.249702	0.499702	0.19981	0.447	1.0
0.159610	039951 0	0.13764	0.371	0 .8
0.089763	0.299604	0.08237	0.287	0 .6
0.039951	0.199877	0.03842	0.196	0 .4
0.010100	0.100504	0.01000	0.100	0 .2
0.002506	0.050063	0.0025	0.05	0.1
0	0	0	0	0

***Notice the relationship between *d*, r, and** r^2

Table (3): Effect size, Sample size &Power (Alpha=0.05; Power=0.95).*

Effect size	Delta	Critical t	Total sample size	Actual power
0.001	3.605	1.960	51978840	0.950
0.100	3.606	1.960	5200	0.950
0.200	3.608	1.962	1302	0.950
0.300	3.613	1.964	580	0.950
0.400	3.622	1.967	328	0.951
0.500	3.623	1.971	210	0.950
0.600	3.650	1.976	148	0.952
0.700	3.671	1.982	110	0.953
0.800	3.666	1.989	84	0.952

0.900	**3.711**	**1.997**	**68**	**0.955**
10.00	**10.000**	**4.303**	**4**	**0.993**

Footnotes:*Power depends on the effect size, the sample size and the significance level.

Table (4): Calculate d and r using means and st.ds for two groups.

Group I				**Group II**			
M1	**SD1**	**Cohen's d**	**Effect size r**	**M2**	**SD2**	**Cohen's d**	**Effect size r**
1	**1**	**0**	**0**	**1**	**1**	**0**	**0**
2	**5**	**0.505-**	**0.245-**	**6**	**10**	**0.505-**	**0.245-**
5	**10**	**0.632-**	**0.302-**	**10**	**5**	**0.632-**	**0.302-**
5	**10**	**0.5**	**0.243**	**0**	**10**	**0.5**	**0.243**
15	**50**	**0.1-**	**0.049-**	**20**	**50**	**0.1-**	**0.049-**
20	**50**	**0**	**0**	**2**	**10**	**0**	**0.5**
50	**100**	**0.380**	**0.186**	**20**	**50**	**0.380**	**0.186**
50	**100**	**-0.280**	**0.139-**	**50**	**100**	**-0.280**	**0.139-**

Note :d and r are positive if the mean difference is in the predicted direction.

$$\text{Cohen's d} = \frac{m_1 - m_2}{\sigma_{\text{pooled}}} \quad , \sigma_{\text{pooled}} = (\sigma_1^2 + \sigma_2^2)/2$$

Table (5): Calculate d and r using t values and degrees of freedom.

T value	**D f**	**Cohen's d**	**Effect size r**
1	**1**	**2**	**0.7071**
1.5	**2**	**2.1213**	**0.7276**
2.0	**5**	**1.7888**	**0.6666**
2.0	**10**	**1.2649**	**0.5345**
2.5	**30**	**0.9128**	**0.4152**
3.0	**30**	**1.0954**	**0.4803**
3.0	**50**	**0.8485**	**03905**

$$\text{Cohen's d} = \frac{2t}{dt^2} \quad , \quad r = \left(\frac{t^2}{t^2 + df}\right)^2$$

Note: d and r are positive if the mean difference is in the predicted direction.

Dr. habib Ahmed Elsayir received his Ph.D degree in Statistics from Omdurman Islamic University, Sudan in 2001. He was appointed manager of Omdurman Islamic University Branch at AlDaein 2002-2005.Now he is assistant prof in the Dept. of Mathematics ,Alqunfudha University college, University of um Al Qura,Saudi Arabia