East Carolina University
Department of Psychology

Nonparametric Effect Size Estimators

As you know, the American Psychological Association now emphasizes the reporting of effect size estimates. Since the unit of measure for most criterion variables used in psychological research is arbitrary, standardized effect size estimates, such as Cohen's *d*, η^2, and ω^2 are popular. What is one to use when the analysis has been done with nonparametric methods? This query is addressed in the document "A Call for Greater Use of Nonparametric Statistics," pages 13-15. The authors (Leech & Onwuegbuzie) note that researchers who employ nonparametric analysis generally either do not report effect size estimates or report parametric effect size estimates such as estimated Cohen's *d*. It is, however, known that these effect size estimates are adversely affected by departures from normality and heterogeneity of variances, so they may not be well advised for use with the sort of data which generally motivates a researcher to employ nonparametric analysis.

There are a few nonparametric effect size estimates (see Leech & Onwuegbuzie), but they are not well-known and they are not available in the typical statistical software package.

Remember that nonparametric procedures do not test the same null hypothesis that a parametric *t* test or ANOVA tests. The nonparametric null hypothesis is that the populations be compared are identical in all aspects -- not just in location. If you are willing to assume that the populations do not differ in dispersion or shape, then you can interpret a significant difference as a difference in locations. I shall assume that you are making such assumptions.

With respect to the two independent samples design (comparing means), the following might make sense, but I never seen them done:

- A *d* like estimator calculated by taking the difference in group mean ranks and dividing by the standard deviation of the ranks.
- Another *d* like estimator calculated by taking the difference between the group median scores and dividing by the standard deviation of the scores.
- An eta-squared like estimator calculated as the squared point-biserial correlation between groups and the ranks.

Two Independent Samples

Rank Biserial Correlation. Mike Palij (2015) pointed me to an article by Gene Glass (1966) on the rank biserial correlation. Glass provided these computational formulas for estimating the underlying Spearman rho between two rank variables when one of them has been measured dichotomously. Palij also provided the links to several other resources cited here. Grissom (2015) noted that this statistic is also known as "Cliff's dominance measure (DM)", "delta," and "Cliff's delta." See pages 166-167 in Grissom and Kim (2012) for more details.

These computational formulas follow:

$$rb = \frac{2}{n_0}\left[\bar{Y}_1 - \frac{n+1}{2}\right], \tag{2}$$

$$rb = \frac{2}{n_1}\left[\frac{n+1}{2} - \bar{Y}_0\right]. \tag{3}$$

Formulas (2) and (3) are equivalent. One of the two will probably be very much simpler, depending on the sizes of n_1 and n_0 and the size of the ranks in one group of the dichotomy. (Note that rb is not defined when either n_1 or n_0 is zero.) The computation of rb will be illustrated on the data below:

Scores on Y for

$X = 1$	$X = 0$
10	8
9	6
7	5
4	3
	2
	1

Calculations

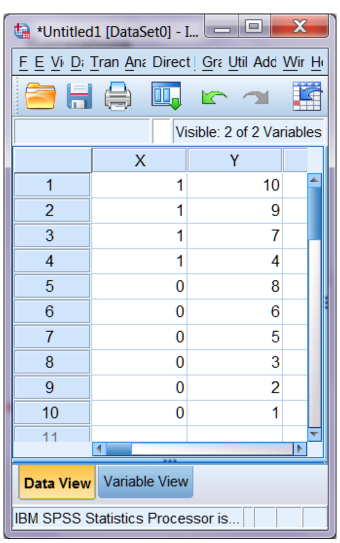
$n_1 = 4,\ n_0 = 6$

$\bar{Y}_1 = 30/4$

$rb = (2/6)(30/4 - 11/2) = 2/3.$

The significance of the difference of rb from zero can be tested by an independent ranks test such as the Mann-Whitney U-test. One

I entered these data into SPSS and computed the U.



Ranks

	X	N	Mean Rank	Sum of Ranks
Y	.00	6	4.17	25.00
	1.00	4	7.50	30.00
	Total	10		

Test Statistics[a]

	Y
Mann-Whitney U	4.000
Wilcoxon W	25.000
Z	-1.706
Asymp. Sig. (2-tailed)	.088
Exact Sig. [2*(1-tailed Sig.)]	.114[b]

Gray and Kinnear (2012) provided this formula for computing the Glass rank biserial correlation coefficient: $\frac{2(M_1 - M_2)}{n_1 + n_2}$, where the means are mean ranks. For Glass' data, that yields 2(7.5 – 4.17) / 10 = .666, the same value Glass reported.

Kerby (2014) provided this formula: $1 - \frac{2U}{n_1 n_2}$. For Glass' data this yields $1 - \frac{2(4)}{6(4)} = .666$, the same value Glass reported.

Normal Approximation *z* to *r*. Pallant (2007) and Field (2013) suggest computing an *r* statistic by taking the *z* that is used for a normal approximation test and dividing it by the square root of the total sample size. Field attributes this to Rosenthal's 1991 version of "Meta-analysis Procedures for Social Research", p 19 (Palij, 2015).

This *r*, using Glass' data, is 1.706/sqrt(10) = .539, which is pretty far from the value of the rank biserial correlation, .666. When I compute the point-biserial correlation here, I found it to be .569, close to the value of the Field/Pallant/Rosenthal coefficient. When I computed the biserial correlation (which assumes normality), I found it to be .722.

Another Data Set. Might the discrepancy between the rank-biserial *r* and the Rosenthal *r* be due to the fact that Glass' data set was very small? I conducted the U test on a sample of 719 students comparing men with women on the height, in inches, of their ideal mates.

Ranks

	Gender	N	Mean Rank	Sum of Ranks
Ideal	Female	539	424.97	229056.50
	Male	180	165.46	29783.50
	Total	719		

Test Statistics[a]

	Ideal
Mann-Whitney U	13493.500
Wilcoxon W	29783.500
Z	-14.638
Asymp. Sig. (2-tailed)	.000

a. Grouping Variable: Gender

Gray and Kinnear $\frac{2(M_1 - M_2)}{n_1 + n_2} = \frac{2(424.97 - 165.46)}{719} = .722$

Kerby $1-\frac{2U}{n_1 n_2}=1-\frac{2(13493.5)}{539(180)}=.722$

Field and Pallant 14.683/sqrt(719) = .548. Nope, sample size does not explain the discrepancy between the Field and Pallant effect size *r* and the Glass rank biserial correlation coefficient.

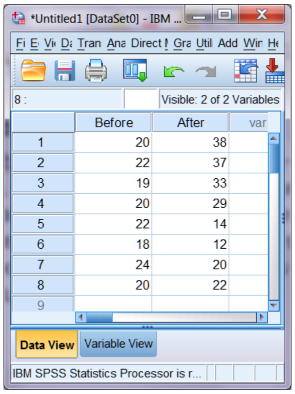
Common Language Effect Size Statistic. This statistic is the probability that a case randomly selected from one group will have a higher score than a case randomly selected from the other group. It has also been called “the probability of superior outcome,” “the probabilistic index,” “probability of concordance,” “measure of stochastic superiority,” $\hat{\theta}$, and AUC, “the Area under ROC curve (Grissom & Kim, 2012; Kraemer, 2006; Newcombe, 2006a, 2006b). When X is dichotomous and Y consists of ranks, an easy way to obtain this statistic is to compute U and then calculate $\hat{p}_{a>b}=\frac{U}{n_a n_b}$. For the height of ideal mate data, $\hat{p}_{a>b}=\frac{13493.5}{539(180)}=.139$, or from the other perspective, 1 - .139 = .861. The expected value of this statistic is, under the null that there is no difference between the groups, .5. The interpretation of the value .87 is that were you randomly to select pairs of scores, one from the male students and one from the female students, 87% of the time the height of the woman’s ideal mate would exceed the height of the man’s ideal mate.

Kraemer (2006) notes that the AUC can be converted to a correlation coefficient this way” CC = 2AUC-1. For the height of ideal mate data, that is 2(.861)-1 = .722 = the value of the rank biserial correlation. For Glass’ data, $\hat{p}_{a>b}=\frac{4}{6(4)}=.1667$, or, from the other perspective, .8333. CC = 2(.833)-1 = .666 = the rank biserial correlation.

Grissom and Kim (2012) cover quite a variety of effect-size estimators for quite a variety of statistical procedures. I highly recommend their book. They also note that if your stats package does not compute U, but rather computes the Wilcoxon Rank Sum Statistic, you can get $U=W-\frac{n_s(n_s+1)}{2}$, where n_s is the smaller of n_a and n_b. If there are tied ranks, you may add to *U* one half the number of ties. Newcombe (2006a, 2006b) gives an overview of and evaluation of methods by which one can put a confidence interval on this estimator. You can find, at http://medicine.cf.ac.uk/primary-care-public-health/resources/ , an Excel spreadsheet, GENERALISEDMW.xls , which will calculate for you the value of $\hat{\theta}$, a confidence interval for $\hat{\theta}$, and the corresponding value of estimated Cohen’s d (delta) with its confidence interval. The user only need enter the sample sizes and the value of the Mann-Whitney *U* statistic.

Correlated Samples

Matched-Pairs Rank-Biserial *r*. Kerby (2014) presents a simple method for obtaining the matched-pairs rank-biserial correlation. First, conduct the Wilcoxon signed ranks test. The output will include the sum of negative ranks and the sum of positive ranks. Add these two sums to get the total sum of ranks. Then divide each of these by the total sum of ranks to get a proportion. The difference between these proportions is the matched-pairs rank-biserial correlation. Here I shall use the data used by Kerby to illustrate the procedure



	Before	After
1	20	38
2	22	37
3	19	33
4	20	29
5	22	14
6	18	12
7	24	20
8	20	22

Ranks

		N	Mean Rank	Sum of Ranks
After - Before	Negative Ranks	3[a]	3.00	9.00
	Positive Ranks	5[b]	5.40	27.00
	Ties	0[c]		
	Total	8		

Test Statistics[a]

	After - Before
Z	-1.260[b]
Asymp. Sig. (2-tailed)	.208
Exact Sig. (2-tailed)	.250
Exact Sig. (1-tailed)	.125
Point Probability	.027

Total sum of ranks = 9 + 27 = 36.

Proportion for positive ranks = 27/36 = .75.

Proportion for negative ranks = 9/36 = .25.

Matched-pairs rank-biserial correlation = .75 - .25 = .5.

Normal Approximation *z* to *r*. Pallant (2007) recommended taking the *z* score normal approximation statistic and dividing it by the square root of *N*, where *N* is the total number of scores (twice the number of cases). For Kerby's data, that is 1.26 / sqrt(16) = .315.

For the exact test, the test statistic, *T*, is the smaller of the two sums of ranks. For these data, *T* = 9, and the .05 lower-tailed critical value for nondirectional hypotheses is 4. Since the obtained *T* is not lower than the critical value, the null is retained. Note that SPSS provides the exact *p*, .25.

Common Language Effect Size Statistic. The common language effect size statistic for data like these is quite simply the proportion of cases (matched pairs) for which the score in the one condition is greater than the score in the other condition. If there are ties, one can simply discard the ties (reducing *n*) or add to the numerator one half the number of ties. For the data above, The CL statistic has value 5/8 = .625. If you were randomly to select one case, the probability that the after score would exceed the before score is 62.5%.

You can find SAS code for computing two nonparametric effect size estimates in the document "Robust Effect Size Estimates and Meta-Analytic Tests of Homogeneity" (Hogarty & Kromrey, SAS Users Group International Conference, Indianapolis, April, 2000).

I posted a query about nonparametric effect size estimators on EDSTAT-L and got a few responses, which I provide here.

Leech (2002) suggested to report nonparametric effect size indices, such as Vargha & Delaney's A or Cliff's d. (Leech (2002). A Call for Greater Use of Nonparametric Statistics. Paper presented at the Annual Meeting of the Mid-South Educational Research Association, Chattanooga, TN, November 6-8.)

John Mark, Regions University.

See the chapter titled "Effect sizes for ordinal categorical variables" in Grissom and Kim (2005). Effect sizes for research. Lawrence Erlbaum.

dale

If you find any good Internet resources on this topic, please do pass them on to me so I can include them here. Thanks.

References

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. Sage.

Glass, G. V. (1966). Note on rank biserial correlation. *Educational and Psychological Measurement, 26*, 623-631.

Gray, C. D., & Kinnear, P. R. (2012). *IBM SPSS statistics 19 made simple*. Psychology Press.

Grissom, R. J. (2015). *Nonparametric effect size estimators*. Post in EDSTAT-L (listserv), 16-June-2015.

Grissom, R. J., & Kim, J. J. (2012). *Effect sizes for research: Univariate and multivariate applications* (2nd edition). New York, NY: Routledge.

Kerby, D. S. (2014). The simple difference formula: an approach to teaching nonparametric correlation. *Innovative Teaching*, *3*(1), Article-1.

Kraemer, H. C. (2006). Correlation coefficients in medical research: from product moment correlation to the odds ratio. *Statistical Methods in Medical Research*, *15*: 525–545

Newcombe, R. G. (2006a). Confidence intervals for an effect size measure based on the Mann–Whitney statistic. Part 1: General issues and tail-area-based methods. *Statistics in Medicine*, *25*, 559 – 573.

Newcombe, R. G. (2006b). Confidence intervals for an effect size measure based on the Mann–Whitney statistic. Part 2: Asymptotic methods and evaluation. *Statistics in Medicine*, *25*, 559 – 573.

Palij, M. (2015). *Nonparametric effect size estimator*. Post in EDSTAT-L (listserv), 13-June-2015.

Pallant, J. (2007). *SPSS Survival Manual 3rd Edition*. New York, NY: McGraw Hill Open University Press.

- *CL*: The Common Language Effect Size Statistic
- Return to Dr. Wuensch's Statistics Lessons Page

This document most recently revised on 16-June-2015.

Karl L. Wuensch, June, 2015