Psychological Reports: Measures & Statistics
2014, 115, 3, 741-747.

A NOTE ON COHEN'S OVERLAPPING PROPORTIONS OF NORMAL DISTRIBUTIONS[1]

JAMES W. GRICE

Department of Psychology
Oklahoma State University at Stillwater

PAUL T. BARRETT

Cognadev UK Ltd.

Summary.—Social scientists are often interested in computing the proportion of overlap and nonoverlap between two normal distributions that are separated by some magnitude. In his popular book, *Statistical Power Analysis for the Behavioral Sciences* (1988, 2nd ed.), Jacob Cohen provided a table (Table 2.2.1) for determining such proportions from common values of separation. Unfortunately, Cohen's proportions are inconsistent with his explication of the popular index of effect size, *d*; and his proportions are underestimates of distributional overlap and overestimates of nonoverlap. The authors explain how Cohen derived his values and then provide a revised, corrected table of proportions that also match values presented elsewhere.

Jacob Cohen's *d* is a popular and simple index of effect size. In his book, *Statistical Power Analysis for the Behavioral Sciences* (1988, 2nd ed.), Cohen discusses *d* in the context of *t* tests for means and defines it as

$$d = \frac{m_A - m_B}{\sigma},$$

where m_A and m_B are population means expressed in the original units of observation, and σ is the common population standard deviation. Although defined as population values, the means and standard deviation in the formula are often estimated from randomly drawn samples or from observations randomly assigned to two groups.

As is clear in the formula, *d* is like the common *z* score because it represents the standardized difference between two means. Values near zero consequently indicate small differences or effect sizes. Similarly, it is well known that if *z* scores are assumed to be normally distributed, they can be converted to percentiles; e.g., a *z* score equal to 1.64 represents the ~95th percentile of the normal distribution (i.e., ~95% of the values are equal to or less than 1.64). In his 1988 book, Cohen routinely assumes the two means, m_A and m_B, are drawn from normal distributions (due to the Central Limit Theorem), and he uses this assumption to derive proportional or percentile indices to aid in the interpretation of *d*. For example:

> "An experimental psychologist designs a study to appraise the effect of opportunity to explore a maze without reward on subse-

[1]Address correspondence to James W. Grice, Department of Psychology, Oklahoma State University at Stillwater, Stillwater, OK 74078 or e-mail (james.grice@okstate.edu).

DOI 10.2466/03.PR0.115c29z4 ISSN 0033-2941

> quent maze learning in rats. Random samples of 30 cases each are drawn from the available supply and assigned to an experimental (E) group which is given an exploratory period and a control (C) group, which is not. Following this, the 60 rats are tested and the number of trials needed to reach a criterion of two successive errorless runs is determined. The (nondirectional) null hypothesis is $|m_E$ and $m_C| = 0$. She anticipates that the ES [effect size] would be such that the highest 60% of one population would exceed the lowest 60% of the other…" (p. 40).

In this example and others in his text, Cohen shifts the understanding of means from *per se* averages to actual scores that can be drawn from normal distributions. The formula for d makes this shift clear as well, as the difference between two means (essentially treated as raw scores) is divided by the common variability of the raw scores. The 60% computed by Cohen in the example therefore reflects an entirely empirical and thoroughly frequentist understanding of d. As a consequence of this position, the computation he provides for one of his proportional indices is incorrect.

Cohen writes, "If we maintain the assumption that the populations being compared are normal and with equal variability, and conceive them further as equally numerous, it is possible to define measures of nonoverlap (**U**) associated with d which are intuitively compelling and meaningful" (p. 21). Again, by writing "equally numerous" Cohen is indicating his frequentist understanding of d, and the goal of determining **U** (hereforth, $\mathbf{U}_1$) is to aid with its interpretation. He goes on to present examples demonstrating the monotonic relationship between d and the percentage of nonoverlap between two normal distributions, reporting the results in Table 2.2.1 (p. 22). When $d = 0.1$, for instance, Cohen writes, "$\mathbf{U}_1$ here equals 7.7%, i.e., 7.7% of the area covered by both populations combined is not overlapped" (p. 21) and when $d = 2$, "$\mathbf{U}_1$ then equals 81.1%, the amount of combined area not shared by the two population distributions" (p. 22).

These descriptions, however, do not agree with the meaning of the proportions as defined in Cohen's first frequentist statement. Consider when d equals 2, and refer to the top part of Fig. 1 below. As can be seen, the authors assume both distributions, labeled "control" and "experimental," are normal with standard deviations equal to unity. Their means are 0 and 2, respectively; thus, $d = [(2 - 0)/1] = 2$. The "intersection point" is $(2 - 0)/2 = 1$. The proportion of the experimental distribution that equals or exceeds 1 is 0.8413, and the proportion that equals or is less than 1 is 0.1587. These values can be found in any standard normal z table, and they are shown in the bottom part of Fig. 1. Incorporating the latter proportion (0.1587) in the overlapped distributions in Fig. 2, it can readily be seen

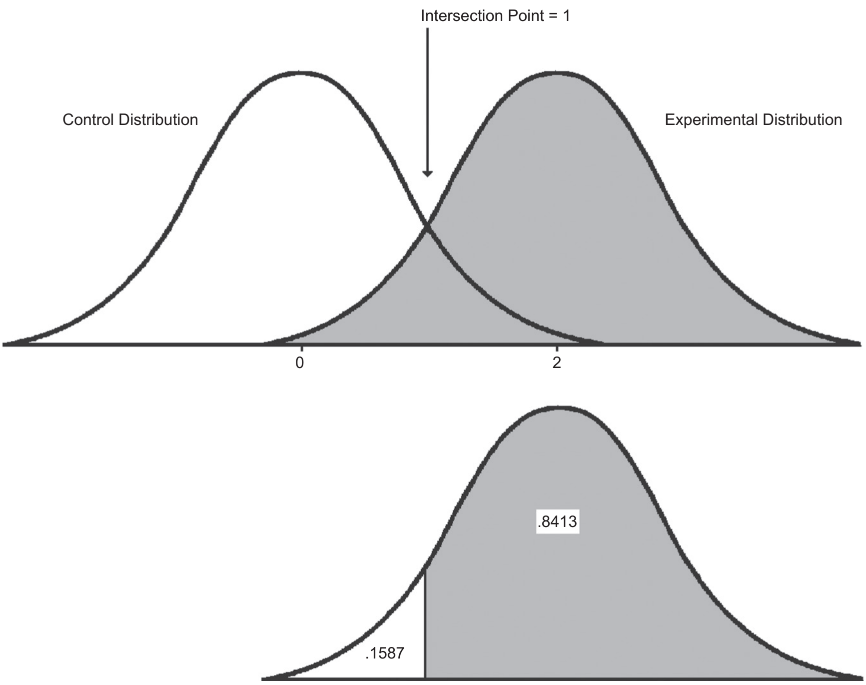
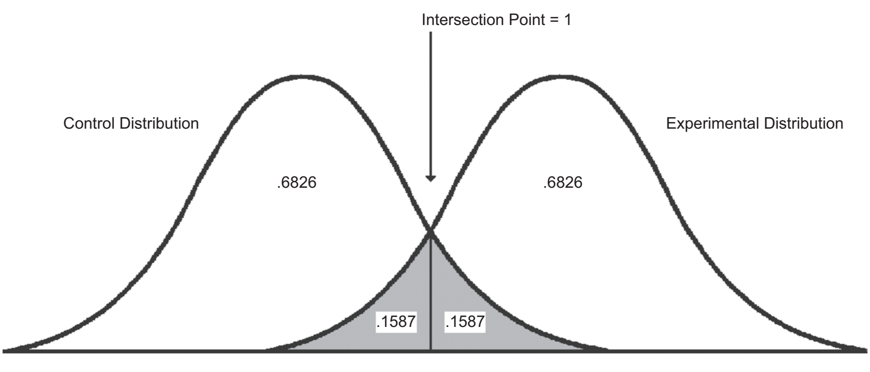


FIG. 1. Overlapping control and experimental distributions, and experimental distribution with shaded proportions



Total Area of Experimental Distribution = .1587 + .1587 + .6826 = 1
Total Area of Control Distribution = .6826 + .1587 + .1587 = 1
Total Area of both Distributions = 2
Overall proportion of Nonoverlap = (.6826 + .6826) / 2 = .6826

FIG. 2. Overlapping distributions with shaded areas

that the proportion of nonoverlap (non-shaded area) for the experimental distribution is 0.6826. Because the two distributions are identical in shape and are "equally numerous," as Cohen states, the overall proportion of nonoverlap is also 0.6826; computed as (0.6826 + 0.6826)/2 = 0.6826, where 2 equals the total area of both distributions. This value does not match Cohen's 0.811 (81.1%), but it can be approximated from a chart provided by Linacre (1996; http://www.rasch.org/rmt/rmt101r.htm).

How did Cohen obtain 81.1%? The key is to realize that Cohen subtly shifted from thinking about frequencies to thinking about areas. Note his first statement, "If we maintain the assumption that the populations being compared are normal and with equal variability, and conceive them further as equally numerous…" Clearly, he is speaking about frequencies of two populations of observations that are normally distributed. When speaking of actual $\mathbf{U}_1$ values, however, he shifts to speaking about areas; for instance, "$\mathbf{U}_1$ here equals 7.7%, i.e., 7.7% of the *area covered by both populations combined* is not overlapped" (emphasis added). Focusing on area, as depicted in Fig. 3, we can compute the total area as 0.6826 + 0.3174 + 0.6826 = 1.6826. The combined nonoverlapping area can then be computed as (0.6826 + 0.6826)/1.6826 = 0.8114, or 81.1%; or, as Cohen reports on page 23,

$$U_1 = \frac{2U_2 - 1}{U_2},$$

where $\mathbf{U}_2$ equals the area of the experimental distribution above the intersection point [e.g., 0.8413; $\mathbf{U}_1 = ((1.6826 - 1)/0.8413) = 0.8114$]. All other values in Cohen's Table 2.2.1 can be computed in similar fashion. Consequently, they do not express the proportions of overall nonoverlapping scores, although these are the values Cohen desired given his empirical, frequentist explication of *d*.

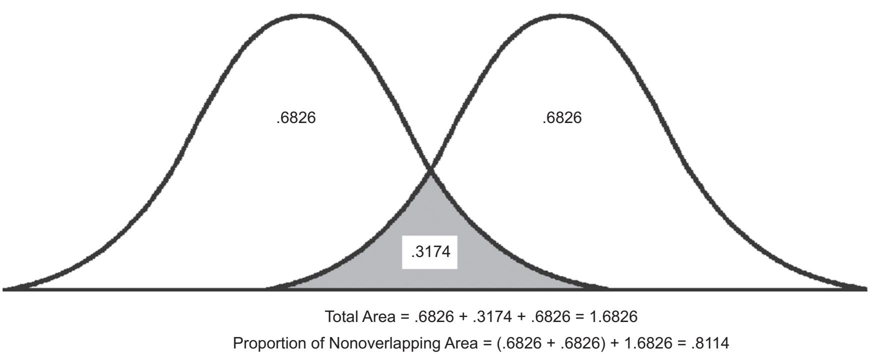


FIG. 3. Overlapping distributions with the focus on combined area

Table 1 presents a revised version of Cohen's Table 2.2.1 with the overall percent nonoverlap and overlap for different values of *d*. The values in this revised table were computed using the reasoning above and were checked against approximate values obtained from Linacre's chart cited

TABLE 1
OVERALL PERCENT OVERLAP FOR DIFFERENT VALUES OF *D* FOR TWO NORMAL DISTRIBUTIONS WITH EQUAL STANDARD DEVIATIONS

d	Cohen's $\mathbf{U}_1$	Overall Percent Nonoverlap	Overall Percent Overlap
0.0	0.0	0.00	100.00
0.1	7.7	3.99	96.01
0.2	14.7	7.97	92.03
0.3	21.3	11.92	88.08
0.4	27.4	15.85	84.15
0.5	33.0	19.74	80.26
0.6	38.2	23.58	76.42
0.7	43.0	27.37	72.63
0.8	47.4	31.08	68.92
0.9	51.6	34.73	65.27
1.0	55.4	38.29	61.71
1.1	58.9	41.77	58.23
1.2	62.2	45.15	54.85
1.3	65.3	48.43	51.57
1.4	68.1	51.61	48.39
1.5	70.7	54.67	45.33
1.6	73.1	57.63	42.37
1.7	75.4	60.47	39.53
1.8	77.4	63.19	36.81
1.9	79.4	65.79	34.21
2.0	81.1	68.27	31.73
2.2	84.3	72.87	27.13
2.4	87.0	76.99	23.01
2.6	89.3	80.64	19.36
2.8	91.2	83.85	16.15
3.0	92.8	86.64	13.36
3.2	94.2	89.04	10.96
3.4	95.3	91.09	8.91
3.6	96.3	92.81	7.19
3.8	97.0	94.26	5.74
4.0	97.7	95.45	4.55

above. Comparing the values in Table 1 to those in Cohen's Table 2.2.1 reveals that the general consequence of miscomputing $\mathbf{U}_1$ is an inflation of the magnitude of nonoverlap and a deflation of the magnitude of overlap. Interpretations of effect sizes will necessarily be inflated or deflated as well. For instance, David Howell (*Statistical Methods for Psychology*, 6th ed., 2007) reports Cohen's proportions (0.85, 0.67, and 0.53) for three levels of *d* (0.2, 0.5, 0.8, respectively) and describes them as "the degree to which the two distributions…overlap" (p. 218) in a frequentist context. The correct proportions of overlap are 0.92, 0.80, 0.69, respectively. As another example, consider this lengthier quote from Levine and Parkinson's (1994) *Experimental Methods in Psychology*:

> "As was stated previously, when $d'=0$ there is no detection. When $d'=0$ there is no distance between the means of the signal-plus-noise and noise distributions. Another way of saying this is that the two distributions are 100% overlapped; there is 0% nonoverlap. As *d*' takes on values greater than zero, some portion of the area covered by both distributions combined will not be overlapped. The percentage nonoverlap provides us with a way of thinking about discriminability and values of *d*'. Consider a *d*' value of 0.50. The means of the signal-plus-noise and noise distributions are separated by 0.50 standard deviations. This means that a portion of the area covered by both distributions does not overlap. Cohen (1988, Table 2.2.1) listed the percentage nonoverlap between two normal distributions separated by distances from 0 to 4 standard deviation units. Referring to his table, we find that 33% of the combined area is nonoverlapped when $d'=0.50$. That is, 33% of the area covered by the signal-plus-noise and noise distributions combined is either noise or signal plus noise, but not both." (p. 232)

Here the authors use *d*' to refer to the standardized difference between means (*d*). They are also clearly thinking about frequencies of observations (values) and interpreting Cohen's $\mathbf{U}_1$ as the overall proportion of nonoverlap in a signal-detection paradigm, *but it does not represent this proportion*. Instead, when *d* is equal to 0.50, the actual overall proportion of nonoverlap between two normally distributed sets of values with equal standard deviations is 0.20 (0.1974), and the overall proportion of overlap is 0.80 (0.8026).

The examples from Cohen's 1988 text are similar, and each invokes distributions of hypothesized, normally distributed scores centered at their respective means. It is the *proportion of nonoverlapping values* between these distributions that Cohen was seeking to compute, not the proportion of unique area between the distributions. Empirical researchers will routinely be seeking the proportion of nonoverlapping scores as well, and for this reason the authors recommend the use of values from Table 1 rather than the values from Cohen's Table 2.2.1.

REFERENCES

COHEN, J. (1988) *Statistical power analysis for the behavioral sciences*. (2nd ed.) Hillsdale, NJ: Erlbaum.

HOWELL, D. C. (2007) *Statistical methods for psychology*. (6th ed.) Belmont, CA: Thomson.

LEVINE, G., & PARKINSON, S. (1994) *Experimental methods in psychology*. Hillsdale, NJ: Erlbaum.

LINACRE, J. M. (1996) Overlapping normal distributions. *Rasch Measurement Transactions*, 10(1), 487-488.

Accepted October 24, 2014.