Using the extreme groups strategy when measures are not normally distributed
— Source link ⬈

Robert L. Fowler

Institutions: University of South Florida

Related papers:

- The use of extreme groups to test for the presence of a relationship

- Modeling the variability of rankings

- ANCOVA based on comparing a robust measure of location at empirically determined design points

- The influence of parent population distribution on d2 values

- A test for comparing large sets of tau values

Using the Extreme Groups Strategy When Measures Are Not Normally Distributed

Robert L. Fowler
University of South Florida

The extreme groups research strategy is a two-stage measurement procedure that may be employed when it is relatively simple and inexpensive to obtain data on a psychological variable (X) in the first stage of investigation, but it is quite complex and expensive to measure subsequently a second variable (Y). This strategy is related to the selection of upper and lower groups for item discrimination analysis (Kelley, 1939) and to the treatments × blocks design in which participants are first "blocked" on the X variable and then only the extreme (highest and lowest means) blocks are compared on the Y variable, usually by a t test or an analysis of variance. Feldt (1961) showed analytically that if the population correlation coefficient between X and Y is $\rho = .10$, the power of the t test is maximized if each extreme group consists of 27% of the population tested on the X variable. However, Feldt's derivation assumes that the X and Y variables are normally distributed. The present study employed a monte carlo simulation to explore the question of how to optimize power in the extreme groups strategy when sampling from non-normal distributions. The results showed that the optimum percent for the extreme group selection was approximately the same for all population shapes except for the extremely platykurtic (uniform) distribution. The power of the extreme groups strategy under conditions of normality was compared to the power of other research strategies, and an extension of the extreme groups approach was developed and applied in an example. *Index terms: construct validation; extreme-group design; monte carlo technique; non-normal distributions; statistical power; upper-lower index.*

The extreme groups research strategy is a two-stage procedure that sometimes is employed when it is relatively simple and inexpensive to measure a psychological variable (X) in the first stage of investigation, but it is quite complex and expensive to obtain data subsequently on a second variable (Y). This situation often exists in the early stages of validating a concept such as was done, for example, by Taylor (1951) with the Manifest Anxiety Scale and by Glass (1977) with the Type A Behavior Pattern. This approach is related to the treatments × blocks experimental design (Myers, 1979) in which participants are first "blocked" on the X variable and then only the extreme (highest and lowest means) blocks are used in collecting data on the Y variable. The means of the two blocks then are compared, usually by a t test or an analysis of variance (ANOVA).

Feldt (1961) showed analytically that if the population correlation coefficient between X and Y is $\rho = .10$, the power of the t test is maximized if each extreme group consists of 27% of the population tested on the X variable. As ρ increases, the optimum percent decreases slightly; for example, at $\rho = .8$, power is greatest when 23% of each tail of the X distribution is used. This is similar to the result obtained by Kelley (1939) in the selection of upper and lower groups for item discrimination analysis, in which the point-biserial correlation coefficient (algebraically convertible to t for statistical significance testing) is often used as an index of discriminability. Cureton (1957), in a more formal mathematical proof of the same relationship, assumed not only bivariate normality, but also independence of the tails of the Y distribution (the equivalent of $\rho = 0$). Subsequently, D'Agostino

& Cureton (1975) showed that when the assumption of correlated tails is incorporated into the analysis, the optimum percent for the upper and lower groups is closer to 21%. In general, during the second phase of the study when information on *Y* is to be obtained, it is a more powerful strategy to use the extreme 25% of each tail of the *X* distribution than to split the *X* distribution at the median and use all of the participants. In essence, greater power is achieved by "throwing away" data—by not using all of the data in the second (*Y*) phase.

All of the derivations mentioned above assumed that the *X* and *Y* variables are normally distributed and linearly related. It is increasingly apparent that real psychological data are rarely normally distributed. For example, Micceri (1989) investigated the distributional characteristics of 440 large-sample achievement and psychometric measures and found all to be non-normal by standard tests of significance. Micceri reported that approximately two-thirds of the 231 achievement test score distributions were at least moderately asymmetric, over 18% of the 125 psychometric measures were exponentially asymmetric, and only 16% of the psychometric distributions were relatively symmetric.

The effects of skewness on the statistical power of the extreme groups strategy may have accounted for some unusual results reported by Kendall (1954) in an early validation of Taylor's Manifest Anxiety Scale. After obtaining anxiety scores on 93 participants, Kendall selected the upper and lower 27% of that distribution (which was positively skewed) to undergo a more extensive clinical evaluation. A *t* test of the difference between the mean clinical ratings of the original two extreme groups failed to reach the .05 level of significance ($p = .083$), but a subsequent *t* test on the upper and lower 13% of the anxiety score distribution yielded a clearly significant outcome ($p = .001$), in spite of the loss of over half of the degrees of freedom.

The primary purpose of the present study was to explore the question of how to select extreme groups in order to maximize the power of *t* to detect the presence of a relationship when the *X* and *Y* score distributions are non-normal. A monte carlo simulation was conducted to examine the power of *t* when samples are drawn from several population distributions, which differed in varying degrees of skewness and/or kurtosis. The study also investigated the effects of the following on the power of *t*: the proportion in the extreme groups (p), the population correlation coefficient (ρ), and the ratio of the number in the upper group (n_u) to the number in the lower group (n_l). Additionally, the power in extreme groups research strategies when *X* and *Y* are normally distributed was examined analytically.

Method

The power of *t* was examined for nine different theoretical population distribution shapes. The first three distributions were symmetric but differed in kurtosis and included: (1) normal (mesokurtic); (2) uniform (platykurtic); and (3) logistic (leptokurtic). The next three distributions were skewed and leptokurtic: (4) truncated or three-quarters normal (slightly skewed); (5) half normal (moderately skewed); and (6) exponential (extremely skewed). The remaining three distributions were symmetric but had long or heavy tails. Sampling from these distributions tends to produce outliers. The remaining three were: (7) normal/uniform; (8) Cauchy; and (9) mixed normal.

The procedure used to generate these distributions by computer was similar to that described by Fowler (1987). Different population shapes were created by using a series of unit uniform pseudorandom numbers, u_i. The methods for generating the various distributions were:

1. Normal or Gaussian (G): $G_i = (-2 \log_e u_1)^{1/2} \cos 2\pi u_2$, after a technique developed by Box & Müller (1958);
2. Uniform (U): $U_i = u_i - .5$;
3. Logistic (L): $L_i = (3)^{1/2} \log_e[u_i/(1 - u_i)]/\pi$;

4. Truncated normal (T): T_i was the G distribution, except that no values of $G_i > .6741891$ were included;
5. Half normal (H): $H_i = -|G_i|$;
6. Exponential (E): $E_i = \log_e u_i + .693147$;
7. Normal/Uniform (G/U): (G_i/U_i);
8. Cauchy (C); $C_i = G_i/G_{i+1}$; and
9. Mixed normal (M): Simulated by sampling from N(0,1) with probability .9, from N(0,9) with probability .095, and from N(0,100) with probability .005. This distribution also is known as a contaminated normal.

The values for each distribution were produced by first drawing N scores from one of these populations. These scores (X) represented the performance of each of the N participants in the first phase of the investigation. Next, for each X value, a second score (Y) was drawn from the same distribution and regressed on X such that the population correlation, ρ_{XY}, was .1, .3, or .5. These values of ρ correspond to Cohen's (1977) small, medium, and large effect sizes, respectively, and represent the range of relationships encountered in most psychological research. Then, the X and Y values in alternate X,Y pairs were exchanged, so that the regression of X on Y would be approximately equal to the regression of Y on X. This step produced identically shaped non-normal populations while maintaining the desired ρ_{XY}. Finally, after ranking the X,Y pairs on the X variable, all possible proportions for the extreme groups, including the case of groups of unequal size, were evaluated by means of estimates of the power parameter for t, which can be expressed as

$$\hat{\Phi} = |\bar{Y}_U - \bar{Y}_L|/[2(S_U^2 + S_L^2)/pN]^{1/2} \tag{1}$$

(Feldt, 1961, Equation 2), where $\bar{Y}$ and S^2 are the sample mean and variance, respectively, of an upper (U) or lower (L) extreme group. This allowed comparisons to be made for a constant value of N without regard to level of significance. Five simulations of $N = 10,000$ using each value of ρ (.1, .3, .5) were run for each of the nine distributions.

Monte Carlo Results

In the empirical phase of the study, all possible proportions for the upper extreme group were combined with all possible proportions for the lower extreme group within the range of $.05 \leq p \leq .50$ in increments of .01. Table 1 presents the results for the extreme groups of equal size case. The values are the points of power maximization (converted to percents) based on the median values for the five simulations conducted with each of the 27 combinations of ρ and distribution shape. The points of optimization at $\rho = .1$ tended to disagree somewhat with those at the higher values of ρ, but the agreement between the empirical and theoretical normal distributions was otherwise quite good. With some of the non-normal populations, notably the E and C distributions, the results of the simulations at $\rho = .1$ often made it difficult to locate a true maximum for the empirical power function because of its flatness over a very wide range of p. This suggests that optimization of power may not be as meaningful in these cases. The most remarkable finding among the groups of equal size is the apparent optimization of power when less than 25% of the extremely platykurtic (uniform) distribution was in each of the extreme groups. This is in marked contrast to Cureton's (1957) analytical proof for this distribution in which the maximum of the function was at 33%. However, Cureton was working with the complete theoretical X distribution rather than a concomitant variable Y, which is related to X only to the extent of ρ.

It was also of interest to determine whether power would be greater in the skewed distributions if more cases were selected from the skewed tail than from the shorter tail. Table 2 shows the median

Table 1
Power Parameter Optimum Percents for Two
Extreme Groups of Equal Size as a Function
of Distribution Shape for Three Levels of ρ

Distribution	$\rho = .1$	$\rho = .3$	$\rho = .5$
Theoretical			
Normal	27	27	26
Empirical			
Normal	32	26	27
Uniform	10	19	19
Logistic	21	29	25
Truncated	26	20	28
Half Normal	31	29	26
Exponential	19	36	32
Normal/Uniform	27	28	32
Cauchy	27	24	32
Mixed Normal	27	25	27

power parameter values for a number of combinations of upper and lower percent extreme groups for the three negatively skewed distributions. For $\rho = .1$, there seemed to be a slight increase in power when a larger percentage was selected from the lower tail (L%) than from the upper tail (U%) of these distributions, but just the opposite appears to be the case when ρ was moderate or large. That is, power tends to be lost when U% > L% in the negatively skewed distributions when $\rho = .1$, compared to the equal percent case. At $\rho = .3$ and $\rho = .5$, however, power decreased when U% < L%. Perhaps the instability of the outcomes obtained in the small effect size situation accounts for this anomaly. This argues against using unequal group sizes to increase power when

Table 2
Median Power Parameters Based on $N = 10,000$ for Various Combinations of Upper (U)
and Lower (L) Percent Extreme Groups in Three Negatively-Skewed Distributions
[Truncated Normal (T), Half Normal (H), Exponential (E)] for Three Levels of ρ

Group Percent		$\rho = .1$			$\rho = .3$			$\rho = .5$		
U%	L%	T	H	E	T	H	E	T	H	E
10	20	10.38	7.83	11.39	24.51	18.65	23.28	40.81	28.77	34.51
15	15	9.78	8.21	11.01	27.28	21.84	27.22	46.42	33.50	39.12
20	10	9.18	6.40	9.94	26.65	22.62	28.92	47.80	36.51	42.96
10	30	9.91	7.59	11.51	23.42	18.11	22.64	38.72	27.30	32.91
20	20	9.89	7.77	11.32	28.51	23.27	28.88	48.03	36.96	41.46
30	10	8.41	5.61	9.29	25.63	23.99	30.97	47.29	39.50	46.63
10	40	9.78	7.61	11.30	23.01	17.34	21.87	37.01	25.42	31.11
20	30	9.48	7.59	11.65	27.67	23.07	28.34	46.51	35.80	39.91
25	25	8.82	7.08	11.21	28.07	24.90	29.71	48.68	38.55	43.19
30	20	9.25	7.06	10.96	28.09	25.29	31.57	49.01	40.82	45.90
40	10	7.90	5.07	8.39	24.56	23.44	30.39	45.66	40.37	48.26
20	40	9.43	7.72	11.52	27.50	22.37	27.43	44.85	33.76	37.65
30	30	8.89	6.88	11.40	27.42	25.34	31.14	48.21	39.92	44.53
40	20	8.76	6.54	10.16	27.36	25.06	31.64	48.08	42.33	48.13
30	40	8.91	7.07	11.32	27.44	24.75	30.15	46.80	37.82	42.03
35	35	9.14	7.04	10.77	27.51	25.58	30.62	47.79	40.12	44.49
40	30	8.40	6.35	10.67	26.83	25.21	31.37	47.70	41.65	46.91

working with small effects in skewed distributions.

Because the monte carlo simulation revealed minimal effects of non-normality on selecting optimum cut-off points for extreme groups, it appears that the "27% rule" (Cureton, 1957) generally would be applicable in most measurement situations. Consequently, the extreme groups strategy (Feldt, 1961), which is based on assumptions of normality, appears to be a viable approach to increasing power in the validation of psychological measurements. Although there are several versions of the extreme groups strategy that have been derived analytically for normal distributions, precise power comparisons among them as a function of the cut-off point p seem not to have been made in the applied measurement literature.

Power in Extreme Groups Strategies With Normal Data

Feldt (1961) demonstrated the superiority of the extreme groups strategy over the alternative strategy of selecting a random sample from the entire X distribution and correlating these X scores with the corresponding Y scores on the $2pN$ participants selected. Alf & Abrahams (1975) demonstrated analytically that there is a more powerful strategy than either of the strategies considered by Feldt. Their strategy uses the covariance information that is lost when the X scores of the extreme groups are dichotomized to form the two groups for the t test. In essence, a Pearson r is calculated only on the X,Y pairs of the extreme groups. Thus, the Pearson r is superior to the t test (point-biserial r) for the extreme groups strategy, for the same reason that there always will be a loss of power (in a normal distribution) when graduated data are dichotomized artificially (Cohen, 1983).

Alternatively, it might be expected that if the number of blocks in the extreme groups strategy is increased beyond the two most extreme blocks—or if the two extreme groups are each subdivided into greater and lesser extreme parts, thereby restoring some of the "lost" rank categorical information—the power also would increase. This approach is the functional equivalent of converting a point-biserial r for widespread classes (Peters & Van Voorhis, 1940) into a point-polyserial r for widespread classes. The power comparisons of the four strategies can be made for a linear relationship, ρ, between two normally distributed variables, X and Y, by computing ratios of their respective power parameters, Φ.

The power parameter for the Alf & Abrahams (1975) covariance approach (Strategy 1) for extreme groups of equal size may be expressed as:

$$\Phi_1 = \frac{\rho\left[\left(1 + \frac{xz}{p}\right)pN\right]^{1/2}}{(1 - \rho^2)^{1/2}} \ ,$$
(2)

where x is the standard score in a unit normal distribution cutting off the upper p proportion
      of the distribution,
     z is the ordinate at x, and
     p is the proportion in each extreme group.

The power parameters for the Feldt (1961) extreme groups (Strategy 2) and random sample (Strategy 3) techniques are given, respectively, by:

$$\Phi_2 = \frac{\rho \frac{z}{p}}{\left[\frac{1 - \rho^2\left(\frac{z^2}{p^2} - \frac{xz}{p}\right)}{pN}\right]^{1/2}}$$
(3)

and

$$\Phi_3 = \frac{\rho(pN)^{1/2}}{(1 - \rho^2)^{1/2}} \quad .$$

(4)

The power parameter for the proposed alternative (Strategy 4), which involves subdividing the corresponding tail points of the extreme groups into two blocks each (for a total of four blocks), can be derived by extending Feldt's (1961) proof (see Appendix) and can be expressed algebraically as:

$$\Phi_4 = \frac{\rho\left[\dfrac{z_1^2}{p_1} + \dfrac{(z - z_1)^2}{(p - p_1)}\right]^{1/2}}{\left\{\dfrac{p - \rho^2\left[\dfrac{z_1^2}{p_1} + \dfrac{(z - z_1)^2}{(p - p_1)} - xz\right]}{\dfrac{pN}{2}}\right\}^{1/2}} \quad ,$$

(5)

where x_1 is the standard score in a unit normal distribution cutting off the upper p_1 proportion of the distribution,

z_1 is the ordinate at x_1, and

p_1 is the proportion in the most extreme block or subgroup.

The power parameters for Strategies 2, 3, and 4 relative to Strategy 1 were examined. These ratios are shown in Figure 1 for $\rho = .3$, as a function of the proportion in each extreme group. In the case of Strategy 4, it was assumed that the normal curve tails were subdivided into two equal area segments; however, it makes little difference up to a ratio of 5:1 between the two segments. Although the relative power of Strategy 4 does not change much as a function of p (the proportion in an extreme group), it is the least powerful strategy when $p > .2$ (see Figure 1). However, in terms of the effect size index, $f = \Phi/(n)^{1/2}$, where n is the mean number of measurements per block (see Appendix), Strategy 4 is always higher than Strategy 2. Figure 2 depicts the ratio of Φ_4 to Φ_2 and the ratio of f_4 to f_2 for $\rho = .1$ (small effect size) and $\rho = .5$ (large effect size) as a function of the proportion in each extreme group. As the proportion in the extreme groups increases, higher effect

Figure 1
Power of Three Alternative Strategies Relative to the Most Powerful Strategy (Strategy 1)
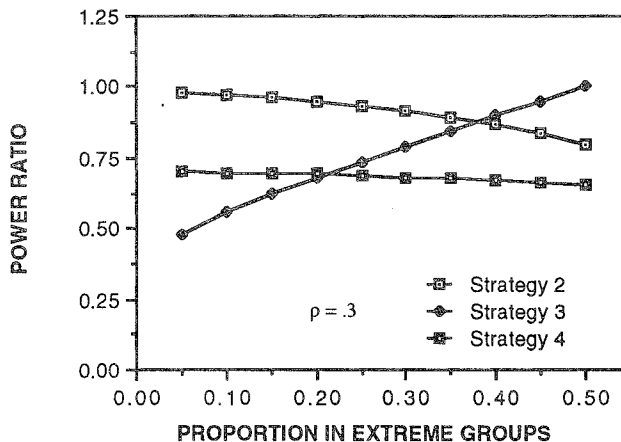
Figure 2
Effect Size (f) and Power (Φ) Ratios of Strategy 4
to Strategy 2 for Small and Large Effect Sizes

a. $\rho = .1$: Small Effect Size

b. $\rho = .5$: Large Effect Size
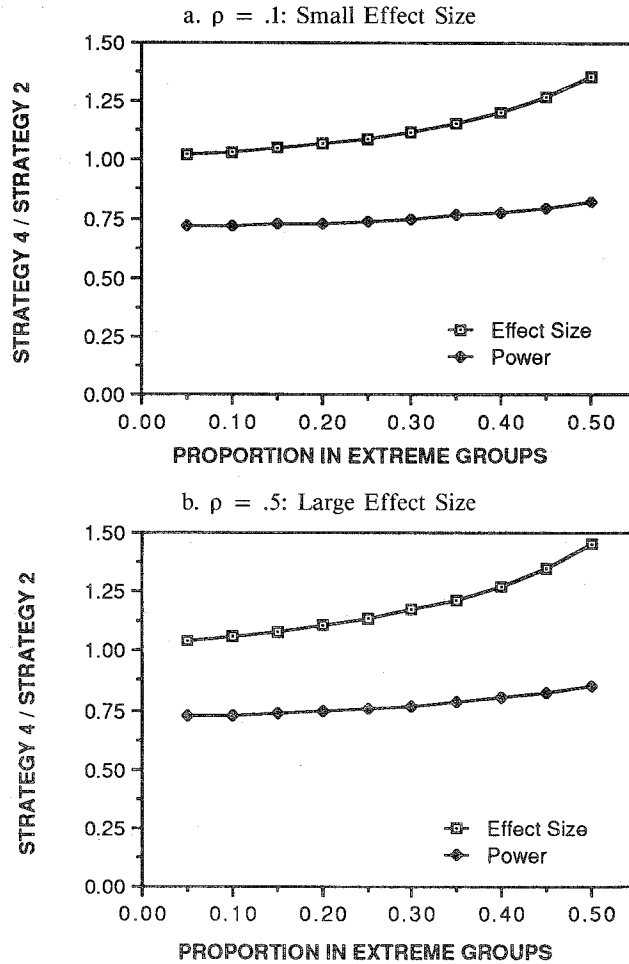
sizes are obtained when two blocks in each extreme group are created, rather than using the traditional two (extreme) group version. Up to $\rho = .5$, the power of Strategy 2 is always superior to that of Strategy 4, regardless of the value of p.

Applying Extreme Groups Strategies

Maximizing power by using the extreme groups approach in test validation research can be an important consideration for organizations, depending on the relative costs of measurement at each of the two stages, as Abrahams & Alf (1978) have shown. Usually, the predictor is less costly to measure than the criterion, but sometimes the reverse is true. Suppose that an organization uses a forced distribution performance appraisal technique and has rated its employees in the following groups with the indicated percentages: (1) superior, 10%; (2) above average, 20%; (3) average, 40%; (4) below average, 20%; (5) Poor, 10%.

By using ANOVA (or point-polyserial r) to compare the mean test score performance of the four groups (excluding the average category), the concurrent postdiction validity statistic will be greater than that for any of the combinations involving only two extreme groups. However, the power to detect a relationship between predictor and criterion will be highest if the top two categories and the bottom two categories are combined into upper and lower extreme groups, respectively.

The power ratios depicted in Figures 1 and 2 are based on parameters reflecting asymptotic results and cannot be used directly to determine power ratios for small and moderate size samples. In order to estimate the power of Strategies 2 and 4 in a particular situation, it is necessary to convert the power parameters to effect sizes by the relationships $f_2 = \Phi_2/(Np)^{1/2}$ and $f_4 = \Phi_4/(Np/2)^{1/2}$, respectively (see Cohen, 1977, pp. 274–284). Suppose that the population correlation between the predictor and criterion measures in the above example is $\rho = .36$. If $p = .3$, $x = .524$, $z = .3478$, $p_1 = .1$, $x_1 = 1.2817$, and $z_1 = .1755$, then $f_2 = .439$ and $f_4 = .473$. The power of Strategies 2 and 4 to detect a relationship may be estimated for any sample size at three conventional levels of significance by using Cohen's (1977, pp. 289–354) power tables. Specifically, for $\alpha = .05$ and $N = 40$, the power of Strategy 2 with upper and lower extreme 30% groups can be found by interpolation in Cohen's tables (p. 311; $u = 1$, $n = pN = 12$, $f = .439$) to be .54, compared to a power for Strategy 4 (p. 315; $u = 3$, $n = pN/2 = 6$, $f = .473$) of .40. It also can be shown that using upper and lower 10% extreme groups (superior vs. poor categories only) would result in a power for Strategy 2 (p. 311, $u = 1$, $n = pN = 4$, $f = .669$) of only .36. If the sample size is increased to $N = 80$, the power estimates of these same three applications of the two strategies increase to .84, .75, and .71, respectively.

In this case, the cost factor may be a consideration if the incumbents, on whom job performance criterion information is already available, must be given time off from work to be measured on the proposed predictor being validated. Thus, eliminating the middle 40% of the criterion distribution from the analysis not only conserves employee production time, but also can result in increased power to detect the presence of a predictor-criterion relationship.

Appendix

The quantity $\rho/(1 - \rho^2)^{1/2}$ that appears in each of the power equations given by Feldt (1961, Equations 7 and 11) and by Alf & Abrahams (1975, Equation 12) is equivalent to the effect size index f, in the F test on means in ANOVA (Cohen, 1977, p. 284). That is,

$$f = \left(\frac{\rho^2}{1 - \rho^2}\right)^{1/2} \quad . \tag{6}$$

In turn, f^2 is simply the ratio of the between-groups sum of squares (SS_b) to the within-groups sum of squares (SS_w) in a one-way ANOVA (Cohen, 1977, pp. 281–283) or

$$f^2 = \frac{SS_b}{SS_w} \quad . \tag{7}$$

The derivation of the power parameter Φ for Strategy 4 can be accomplished by extending the proof for two extreme groups of equal size (Feldt, 1961, pp. 308–310) to the case of four extreme subgroups (with the corresponding upper and lower subgroups of equal size) and expressing it in ANOVA terms. Assume that X and Y are each N(0,1) and are correlated at some value ρ.

Let p = the proportion in the upper (or lower) extreme group;
   x = the standard score in a normal distribution cutting off the upper p proportion of

the distribution;
z = the normal curve ordinate at x;
N = the total sample size;
μ_U = the population mean of the upper extreme group; and
μ_L = the population mean of the lower extreme group.

The subscripts 1 and 2 refer to the more extreme subgroup and less extreme subgroup, respectively, and the notations [2] and [4] distinguish between quantities based on two groups and four groups, respectively.

The relationship between the power parameter Φ and the effect size f is given by Cohen (1977, p. 275) as $f = \Phi/(n)^{1/2}$, where $n = pN$ is the number of measurements in each group. Thus, the power parameter for the two extreme groups strategy (Φ_2) given by Feldt (1961, p. 310, Equation 7) can be converted to the effect size as

$$f_{[2]}^2 = \frac{\rho^2 \dfrac{z^2}{p^2}}{1 - \rho^2 \left(\dfrac{z^2}{p^2} - \dfrac{xz}{p} \right)} \quad . \tag{8}$$

Then, for the unit normal curve ($N = 1$) in a one-way ANOVA with two groups of equal size, the SS_b can be determined by

$$SS_{b[2]} = \frac{(p\mu_U)^2}{p} + \frac{(p\mu_L)^2}{p} - \frac{(p\mu_U + p\mu_L)^2}{2p} \quad , \tag{9}$$

which reduces to

$$SS_{b[2]} = 2p\mu_U^2 \quad , \tag{10}$$

when $\mu_U + \mu_L = 0$.

Substituting the appropriate value for the mean of an extreme segment of the normal curve ($\rho z/p$) given by Feldt (1961, pp. 309–310) yields

$$SS_{b[2]} = 2\rho^2 \frac{z^2}{p} \quad , \tag{11}$$

which is $2p$ times the numerator of Equation 8. Therefore, the SS_w must be $2p$ times the denominator of Equation 8 or

$$SS_{w[2]} = 2p \left[1 - \rho^2 \left(\frac{z^2}{p^2} - \frac{xz}{p} \right) \right] \quad . \tag{12}$$

If the two extreme groups are each subdivided (with the corresponding upper and lower subgroups equal size), the SS_b can be determined by

$$SS_{b[4]} = \frac{2(p_1\mu_1)^2}{p_1} + \frac{2(p_2\mu_2)^2}{p_2} - \frac{(p\mu_U + p\mu_L)^2}{2p} \quad , \tag{13}$$

which reduces to

$$SS_{b[4]} = 2(p_1\mu_1^2 + p_2\mu_2^2) \quad . \tag{14}$$

In a one-way ANOVA, the total sum of squares (SS_t) remains the same regardless of the number of groups, so that

$$SS_t = SS_{b[2]} + SS_{w[2]} = SS_{b[4]} + SS_{w[4]} \quad , \tag{15}$$

which means that

$$SS_{w[4]} = SS_{b[2]} + SS_{w[2]} - SS_{b[4]} \tag{16}$$

and

$$f^2_{[4]} = \frac{SS_{b[4]}}{SS_{b[2]} + SS_{w[2]} - SS_{b[4]}} \quad . \tag{17}$$

Substituting the results of Equations 11, 12, and 14 into Equation 17, and noting that $\mu_1 = \rho z_1/p_1$ and $\mu_2 = \rho(z - z_1)/(p - p_1)$, the effect size index becomes

$$f^2_{[4]} = \frac{\dfrac{\rho^2}{p}\left[\dfrac{z_1^2}{p_1} + \dfrac{(z - z_1)^2}{(p - p_1)}\right]}{1 - \dfrac{\rho^2}{p}\left[\dfrac{z_1^2}{p_1} + \dfrac{(z - z_1)^2}{(p - p_1)} - xz\right]} \quad . \tag{18}$$

The power parameter for the F test in ANOVA is given by Cohen (1977, p. 275) as $\Phi = f(n)^{1/2}$, where n is the average sample size per group. In the present case, each of the four subgroups contains an average of $pN/2$ measurements, yielding the final form of the power of Strategy 4 in the case of normal data as

$$\Phi_4 = \frac{\rho\left[\dfrac{z_1^2}{p_1} + \dfrac{(z - z_1)^2}{(p - p_1)}\right]^{1/2}}{\left\{\dfrac{p - \rho^2\left[\dfrac{z_1^2}{p_1} + \dfrac{(z - z_1)^2}{(p - p_1)} - xz\right]}{\dfrac{pN}{2}}\right\}^{1/2}} \quad . \tag{19}$$

References

Abrahams, N. M., & Alf, E. F. (1978). Relative costs and statistical power in the extreme groups approach. *Psychometrika, 43,* 11–17.

Alf, E. F., & Abrahams, N. M. (1975). The use of extreme groups in assessing relationships. *Psychometrika, 40,* 563–572.

Box, G. E. P., & Müller, M. (1958). A note on the generation of random normal deviates. *Annals of Mathematical Statistics, 29,* 610–611.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (rev. ed.). New York: Academic Press.

Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement, 7,* 249–253.

Cureton, E. E. (1957). The upper and lower twenty-seven percent rule. *Psychometrika, 22,* 293–296.

D'Agostino, R. B., & Cureton, E. E. (1975). The 27 percent rule revisited. *Educational and Psychological Measurement, 35,* 47–50.

Feldt, L. S. (1961). The use of extreme groups to test for the presence of a relationship. *Psychometrika, 26,* 307–316.

Fowler, R. L. (1987). Power and robustness in product-moment correlation. *Applied Psychological Measurement, 11,* 419–428.

Glass, D. C. (1977). *Behavior patterns, stress, and coronary disease.* New York: Wiley.

Kelley, T. L. (1939). The selection of upper and lower groups for the validation of test items. *Journal of Educational Psychology, 30,* 17–24.

Kendall, E. (1954). The validity of Taylor's Manifest Anxiety Scale. *Journal of Consulting Psychology, 18,*

429–432.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin, 105,* 156–166.

Myers, J. L. (1979). *Fundamentals of experimental design* (3rd ed.). Boston: Allyn and Bacon.

Peters, C. C., & Van Voorhis, W. R. (1940). *Statistical procedures and their mathematical bases.* New York: McGraw-Hill.

Taylor, J. A. (1951). The relationship of anxiety to the conditioned eyelid response. *Journal of Experimental Psychology, 41,* 81–92.

Author's Address

Send requests for reprints or further information to Robert L. Fowler, University of South Florida, 140 Seventh Avenue South, St. Petersburg FL 33701, U.S.A.