

## **A visual approach to statistical power analysis on the microcomputer**

**MICHAEL BORENSTEIN**

*Hillside Hospital, Division of  
Long Island Jewish Medical Center  
Glen Oaks, New York  
and Biostatistical Programming Associates  
Teaneck, New Jersey*

**JACOB COHEN**

*Biostatistical Programming Associates,  
Teaneck, New Jersey  
and New York University, New York, New York*

**HANNAH R. ROTHSTEIN**

*Biostatistical Programming Associates  
Teaneck, New Jersey  
and Baruch College, City University of New York  
New York, New York*

**SIMCHA POLLACK**

*Hillside Hospital, Division of  
Long Island Jewish Medical Center  
Glen Oaks, New York  
Biostatistical Programming Associates  
Teaneck, New Jersey  
and St. Johns University, Jamaica, New York*

and

**JOHN M. KANE**

*Hillside Hospital, Division of  
Long Island Jewish Medical Center  
Glen Oaks, New York  
and Albert Einstein College of Medicine  
Bronx, New York*

*Computer programs for statistical power analysis typically require the user to provide a series of values and respond by reporting the corresponding power. These programs provide essentially the same functions as a published text, albeit in a more convenient form. In this paper, we describe a program that instead uses innovative graphic techniques to provide insight into the interaction among the factors that determine power. For example, for *t* tests, the means and standard deviations of the two distributions, sample sizes, and alpha are displayed as bar graphs. As the researcher modifies these values, the corresponding values of beta (also displayed as a bar graph) and power are updated and displayed immediately. By displaying all of the factors that are instrumental in determining power, the program ensures*

*that each will be addressed. By allowing the user to determine the impact that any modifications will have on power, the program encourages an appropriate balance between alpha and beta while working within the constraints imposed by a limited sample size. The program also allows the user to generate tables and graphs to document the impact of the various factors on power. In addition, the program enables the user to run on-screen Monte Carlo simulations to demonstrate the importance of adequate statistical power, and as such, it can serve as a unique educational tool.*

In recent years, a number of computer programs have been published that enable researchers to compute statistical power for a variety of procedures (see, e.g., Dunlap, 1981, and Dallal, 1986). Goldstein (1989) has provided a comparative review of 14 programs for computing power. Typically, such programs require the user to provide a series of values (effect size, alpha, sample size); the program then prints the corresponding value of power. A few will also generate tables showing how power will vary as a function of these factors. In a sense, such programs function as electronic books in that they retain the basic approach of published texts, albeit in a more convenient form for microcomputer users.

In this paper, we describe a computer program that instead offers an innovative, comprehensive approach to power analysis for studies that are analyzed by *t* tests. The program features numerous screens for entering data and for displaying the results of calculations. The same data may be entered or modified on a spreadsheet, on a graph showing the amount of overlap between groups, or on a graph showing the relationship between the effect size, the sample size, alpha, and beta. When used in concert with each other, the various screens enable a user to strike a balance between the competing elements that are involved in planning the study. In addition, the program allows the user to run on-screen Monte Carlo simulations to determine power empirically. This feature is a unique educational tool.

### **The Spreadsheet Approach**

Initially, the user accesses the spreadsheet shown in Figure 1 and enters the required values: the means and standard deviations of the two populations, the sample size to be drawn from each population, alpha (the *p* value required for significance), and whether the test of significance is one tailed or two tailed. After entering these values, the user can press a function key to display power, as in Figure 2. Alternatively, however, the user can press a function key to transfer these values to any of the screens described below.

### **Interactive Bar Graphs**

The user may select the screen shown in Figure 3. The parameters that determine power—the distribution of the two populations, the sample size, and alpha—are displayed

---

This research was supported in part by Grants NIMH/SBIR 1-R43-MH-43083 and NIMH MH-41960. The authors also thank the reviewers and Freddy Howard for their comments. H.R.R. is in the Department of Management at Baruch College; S.P. is in the Department of Quantitative Analysis at St. Johns. Correspondence should be addressed to M. Borenstein, Director of Biostatistics, Hillside Hospital, A Division of Long Island Jewish Medical Center, P.O.B. 38, Glen Oaks, NY 11004.

<CURSOR> TO CELL ...                      PRESS <ESC> OR <BACKSPACE> TO ERASE  
 Then use the TOP ROW of the keyboard to enter new value for cell  
 <F3> Set Table Factors                      <F4> View in Context of Effect Size  
 <F8> Switch to Graphics                      <F9> Compute/Proceed                      <F10> EXIT

Mean Population-1	125.000
Mean Population-2	135.000
Std Dev Population-1	15.000
Std Dev Population-2	15.000
N Cases Group 1	30
N Cases Group 2	60
Alpha	0.050
Tails for Alpha	2
Power	

Enter the Mean for Population-1

Figure 1. On the spreadsheet, the user provides the population means and standard deviations, the nominal alpha value, and the sample size; the program will compute the corresponding power. Alternatively, the user may specify the required power and the program will compute the corresponding sample size.

With a sample size of 30 + 60 and alpha 2 tailed set at .050 the observed difference must be about 6.67 in order to be significant

If the population difference is 10.00, then there is a 84% chance <power> that the difference in the sample will be 6.67 or higher

By the same token there is 16% chance <beta> that the difference in the sample will be smaller than 6.67 and NOT significant

GROUP 1	MEAN= 125.000	SD= 15.000	N= 30
GROUP 2	MEAN= 135.000	SD= 15.000	N= 60
COMBINED		SD= 15.000	N= 40

POPULATION DIFFERENCE:	RAW = 10.00	STANDARDIZED <D>	= 0.67
REQUIRED DIFFERENCE:	RAW = 6.67	STANDARDIZED	= 0.44

ALPHA = 0.050                      TAILS = 2                      BETA = 0.16                      POWER = 0.84

<4> EFFECT SIZE                      <7> SPREADSHEET                      <8> GRAPHICS INPUT                      <10> EXIT

Figure 2. The text screen offers a quick display of the projected mean difference, the difference required for statistical significance, and the study's power.

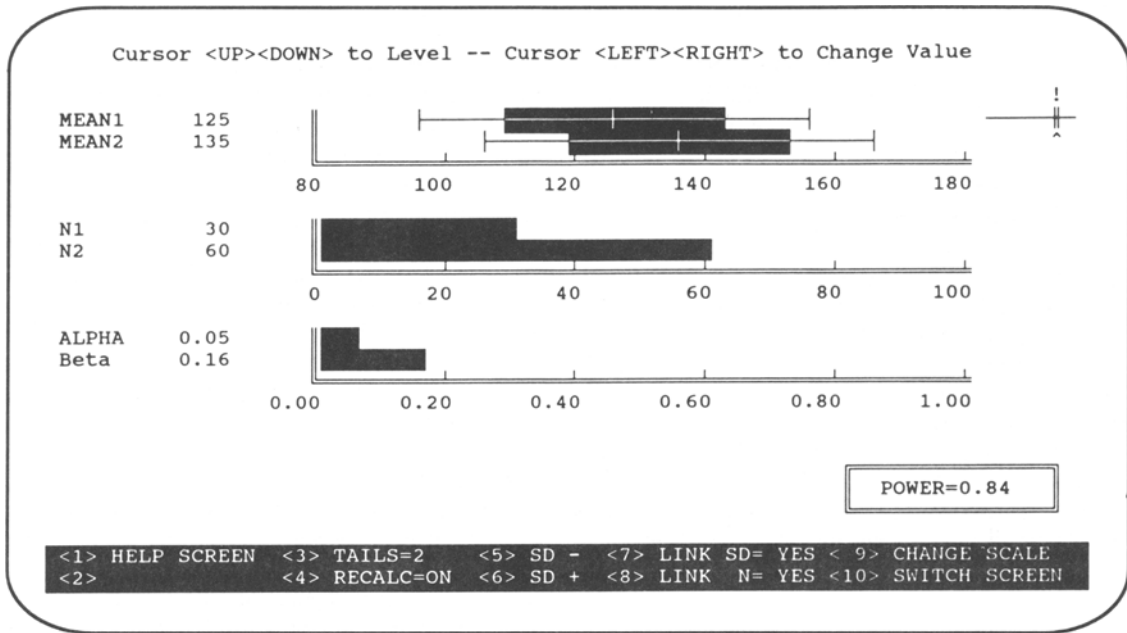


Figure 3. This screen incorporates interactive bar graphs showing the two populations, the N of cases in both groups, and alpha. The researcher uses cursor keys to modify any of these values, and the program responds by updating the bar representing beta.

as bar graphs. The researcher can cursor UP or DOWN to locate any parameter, and then cursor LEFT or RIGHT to modify the value of that parameter. The program immediately updates the value of power.

Population 1 is displayed at the top: The mean is represented by a vertical line, one standard deviation on either side of the mean is indicated by a highlight, and a second standard deviation on either side of the mean is set off by vertical bars. Population 2 is displayed in similar format on the next row. Thus, the degree of overlap between groups becomes a salient feature of the power analysis and the user is able to ensure that the amount of nonoverlap between groups is appropriate. The program allows the user to move the mean of either population toward the left or the right by using cursor keys, and to decrease or increase the standard deviation of either population by using function keys. The impact of these modifications is reflected immediately in the value of power that is displayed at the bottom of the screen.

The next two lines on the main screen (Figure 3) display the number of cases, *N*. The user may cursor LEFT or RIGHT to decrease or increase the sample size. The program can be toggled so that the two groups are of equal size, or the *N*s can be set independently. Again, modifications to the value of *N* are reflected automatically in the display of power.

The next line is used to set the value for alpha. The user may cursor LEFT to move this value toward zero, or RIGHT to move the value toward 1.0. A function key is used to define the alpha value as being either one tailed or two tailed. The value of power is updated immediately in response.

The final bar on this screen displays the value of beta. Whereas the bars described above are under the control of the user, this bar is under the control of the program. As the means, *SD*s, sample sizes, or alpha are modified, the bar representing beta (and the corresponding value displayed for power) moves immediately in response.

Thus, by manipulating the various bars on this screen, the user is able to visualize the impact of study design on power. For example, how would power be affected if the mean of either population were actually a unit lower or higher than the projections? What if the dispersion within populations was somewhat lower or higher than the projected value? What would be the impact on power of adding 20 cases to each group, or of adding 40 cases to the second group? If the researcher decided to use an alpha of .01 rather than .05, what sample size would be required in order to yield an appropriate level of power?

### Effect Size

While the main screen enables the user to visualize quickly the degree of overlap between the two populations, the program also allows the user to evaluate formally the magnitude of the hypothesized effect. Cohen (1977, 1988), noting that the magnitude of effect is a function of the distance between group means relative to the dispersion within a group, developed the index *d*, defined as

$$d = \frac{|M_1 - M_2|}{SD_{within}}$$

(i.e., the standardized mean difference), as a measure of effect size. This *d* will increase as the distance between

means increases and/or the standard deviation within groups decreases. In theory, this index can range in value from zero (indicating no effect) to an indefinitely large positive value. In practice, of course, the upper end of the range varies from one field of study to the next. In the social sciences, for example,  $d$  is rarely as large as 1.00 (Cohen, 1988).

The program allows the user to view the relationship between groups in the context of the effect size index  $d$ . By pressing a function key, the user can invoke the screen shown in Figure 4, where the degree of nonoverlap between the two distributions, shown at the top of the screen, corresponds to the  $d$  of 0.67 displayed at the bottom of the screen. The bar graph indicates that this effect falls in the *medium/large* range by Cohen's (1988) conventions. Again, modifications to the means and  $SD$ s may be made at the top of the screen and are reflected immediately in the bar representing  $d$ . After the two distributions have been modified on this screen, the user would return to the main screen, which would automatically reflect any changes.

### Tables and Graphs

The program also features the ability to generate tables showing how power will vary as a function of effect size, sample size, and alpha (Figure 5) and to display this same information as an on-screen graph (Figure 6). This graph may be used in study planning to identify the point at which increases in power are too small to justify increases in sample size.

### Monte Carlo Simulations

In addition to computing power by formula, the program features the ability to determine power empirically, by simulation. In this procedure, the program creates the two populations that have been specified by the user, displays these populations on the screen, draws a sample of  $N$  cases from each, and then tests the difference between groups for statistical significance (Figure 7). As the process is repeated as many as 10,000 times, the program keeps track of the proportion of samples that yield significance (i.e., the empirical power) and the proportion that do not (i.e., beta). Although the Monte Carlo method here offers no mathematical advantage over the computational method, it provides a compelling vehicle for educational purposes.

### Discussion

Volumes have been written about the textbook application of statistical power analysis: the researcher projects the *true* (i.e., the alternate-hypothetical) effect size delta, selects a value for alpha, and uses a table or a program to find the required sample size. The real world application of power analysis is quite another matter. Feinstein (1975), for example, has argued that the researcher typically begins by selecting a sample size such that (1) the required number of subjects is available and (2) their recruitment can be funded. Alpha is set at the traditional .05 level, a value of delta (the true effect size) is claimed so that power works out to 0.80, and "a suitable rationale is then developed for presentation to the granting agency" (p. 504).

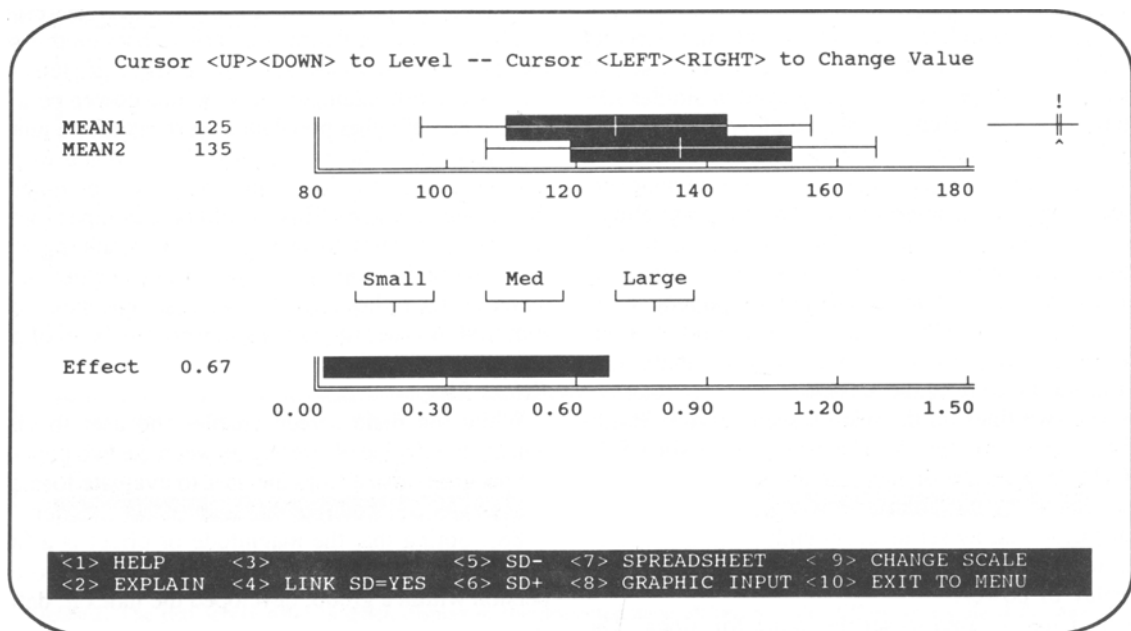


Figure 4. The populations are again displayed at the top, but the bottom half of the screen is used to display the effect size index  $d$ , in the context of Cohen's (1988) suggested conventions for *small*, *medium*, and *large* effect sizes. As cursor keys are used to modify the distributions, the program updates the display of effect size.

MEAN1= 125.0		SD1= 15.0	N1= VARIES	ALPHA= .050		
MEAN2= VARIES		SD2= 15.0	N2= VARIES	TAILS= 2		
N1	N2	ME=	ME=	ME=	ME=	
		132.0	133.0	134.0	135.0	
25	50	0.466	0.572	0.673	0.763	
26	52	0.481	0.589	0.691	0.780	
27	54	0.496	0.606	0.708	0.795	
28	56	0.511	0.622	0.724	0.810	
29	58	0.525	0.637	0.739	0.824	
30	60	0.539	0.653	0.754	0.837	
31	62	0.553	0.667	0.768	0.849	
32	64	0.567	0.681	0.781	0.860	
33	66	0.580	0.695	0.794	0.870	
34	68	0.593	0.708	0.806	0.880	
35	70	0.606	0.721	0.817	0.889	
36	72	0.618	0.733	0.828	0.898	
37	74	0.630	0.745	0.838	0.906	
38	76	0.642	0.757	0.848	0.913	
39	78	0.653	0.768	0.857	0.920	
40	80	0.665	0.778	0.866	0.926	

<4> DISPLAY GRAPH	<5> PRINT FULL TABLE	<6> PRINT SCREEN
CURSOR <UP> OR <DOWN> TO SCROLL TABLE	<10> EXIT TO MENU	

Figure 5. This table shows power as a function of mean difference and sample size. The first column displays power for a mean difference of 125 versus 132; the second column, for a mean difference of 125 versus 133; and so on. The top of the display shows the values of alpha and the standard deviation, which are constant in this example. The program also allows the user to specify that the SDs and/or alpha will vary.

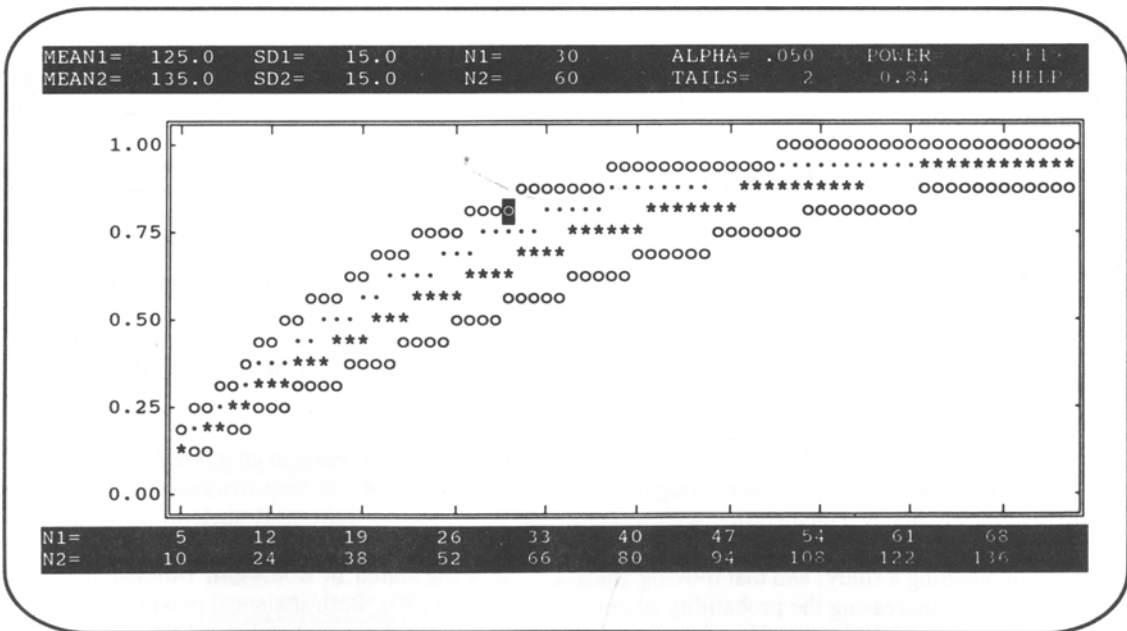


Figure 6. Graph of power as a function of sample size and mean difference. The data correspond to those shown in the table (Figure 5). The four lines on this graph correspond to the four columns in the table (125 vs. 132, 125 vs. 133, etc.). The points along each line correspond to the sample size. Using cursor keys, the user may highlight any point on the graph, and precise values for that point will be displayed in the top margin. In this example, the highlighted point corresponds to means of 125 versus 135, and a sample size of 30 plus 60. The corresponding power is shown as 0.84.

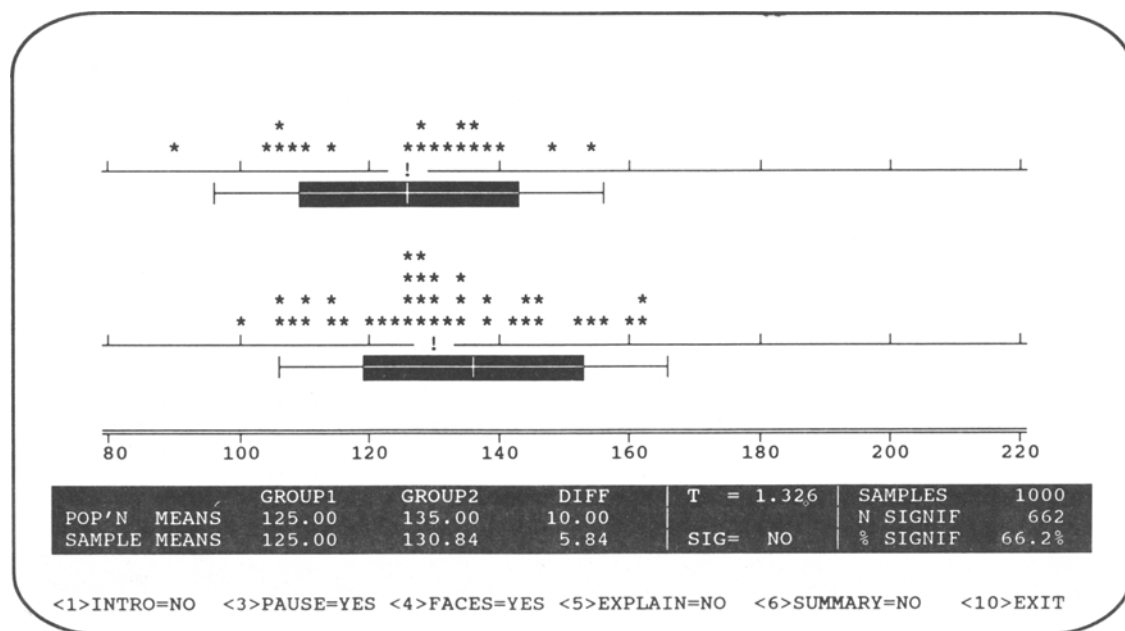


Figure 7. Monte Carlo simulation. The population means of 125 and 135, denoted by the vertical lines, represent a true mean difference of 10 units. The sample means of 125 and 130.8, denoted by the exclamation points, represent an observed mean difference of 5.8 units. The study shown here, which drew 20 cases from the first population and 40 cases from the second, failed to reach the criterion for statistical significance. The tally at the bottom right-hand corner of the screen shows that 66% of the studies have met the criterion for significance, so the empirical power is 0.66.

Our computer program addresses the points made by Feinstein (1975). After the user specifies the means and standard deviations for the two groups, the program presents a graphic display highlighting the amount of overlap and nonoverlap between the groups, thus prompting the user to ensure that the magnitude of the effect is appropriate. As an additional precaution, the program allows the user to view the magnitude of the effect in the context of proposed conventions for *small*, *medium*, and *large* effect sizes. With the magnitude of the effect displayed in these various ways, the researcher is less likely to choose this parameter arbitrarily and would be likely to seek other means for improving power.

The program allows the user to explore these other means in a direct, flexible manner. Modifications to the  $N$  of either or both groups are reflected immediately in the display of power. While alpha is often thought of as fixed at .05, the display of alpha on the same scale as beta serves to highlight the fact that both types of error should be considered in planning a study, and that moving alpha toward the right (i.e., increasing the probability of committing a Type I error) might be considered as a means for moving beta toward the left (i.e., decreasing the probability of committing a Type II error). Cohen (1965) discusses the importance of establishing an appropriate balance between alpha and beta by taking into account the context of the individual study and the real world implications of either type of error.

Tversky and Kahneman (1971) have argued that researchers sometimes fail to appreciate the importance of power because they believe that if a hypothesis is true the study is bound to yield significant results. This mistaken belief is addressed by the program's ability to run Monte Carlo simulations in which the user is able to watch as the samples are drawn. This technique demonstrates in a clear and compelling manner that having a correct hypothesis does not ensure that the study will yield significance, and that adequate power is essential.

Finally, the program enables the researcher to generate tables and graphs so that the various alternatives for increasing power may be carefully documented.

### Related Programs

The reader interested in an innovative approach to computing power for one-way analysis of variance is referred to Borenstein, Cohen, Rothstein, Pollack, and Kane (1990). Similarly, power for hierarchical multiple regression is presented in Rothstein, Borenstein, Cohen, and Pollack (1990). Both statistical power for differences between proportions and statistical power for correlations are discussed in Borenstein and Cohen (1988).

### Algorithms

The formula used to compute power was adapted by Cohen (1988, p. 544) from an algorithm published by Dixon and Massey (1957, p. 253), which is a normal

curve approximation to the noncentral  $t$  distribution. The formula is given by Cohen as

$$z_{1-b} = \frac{d(n-1) \sqrt{2N}}{2(N-1) + 1.21(Z_{1-a} - 1.06)} - z_{1-a},$$

where

$z_{1-b}$  = the percentile of the unit normal curve that gives power,

$z_{1-a}$  = the percentile of the unit normal curve for the significance criterion (for one-tailed tests,  $a = a_1$ , and for two-tailed tests,  $a = a_2/2$ ),

$d$  = the standardized mean difference, and

$n$  = the size of each sample.

The programming function corresponding to this formula is presented in the Appendix. The program draws on algorithms published by Odeh and Evans (1974) to derive the normal curve deviate corresponding to alpha and by Zelen and Severo (1964) to translate the normal curve deviate for  $1-b$  into a value for power. For a two-tailed test, power is computed for both  $d$  and  $-d$ ; the two values are then summed to yield a final value for power.

### Accuracy

The program was checked for accuracy against tables published in Cohen (1988). The values produced by the program are typically identical to those in the book, and never differ by more than 0.01. The program's values were also checked for accuracy against the exact values published by Tiku (1967) and found to agree within 0.01 (Tiku presents values for analysis of variance, and these were converted to  $d$  as described in Borenstein et al., 1990).

### Availability

The program described here will run on any IBM or compatible microcomputer and is available free of charge

from the first author. When requesting the program, indicate the type of video card (IBM monochrome non-graphic, Hercules, CGA, EGA, VGA) and the disk size (5.25 or 3.5 in.).

### REFERENCES

- BORENSTEIN, M., & COHEN, J. (1988). *Statistical power analysis: A computer program*. Hillsdale, NJ: Erlbaum.
- BORENSTEIN, M., COHEN, J., ROTHSTEIN, H. R., POLLACK, S., & KANE, J. M. (1990). Statistical power analysis for one-way analysis of variance: A computer program. *Behavior Research Methods, Instruments, & Computers*, *22*, 271-282.
- COHEN, J. (1965). Some statistical issues in psychological research. In B. B. Wolman (Ed.), *Handbook of clinical psychology* (pp. 95-121). New York: McGraw-Hill.
- COHEN, J. (1977). *Statistical power analysis for the behavioral sciences* (rev. ed.). Hillsdale, NJ: Erlbaum.
- COHEN, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- DALLAL, G. E. (1986). PC-SIZE: A program for sample size determinations. *American Statistician*, *40*, 52.
- DIXON, W. F., & MASSEY, F. J. (1957). *Introduction to statistical analysis* (2nd ed.). New York: McGraw-Hill.
- DUNLAP, W. P. (1981). An interactive FORTRAN IV program for calculating power, sample size, or detectable differences in means. *Behavior Research Methods & Instrumentation*, *13*, 757-759.
- FEINSTEIN, A. R. (1975). The other side of "statistical significance": Alpha, beta, delta, and the calculation of sample size. *Clinical Pharmacology & Therapeutics*, *18*, 491-505.
- GOLDSTEIN, R. (1989). Power and sample size via MS/PC-DOS computers. *American Statistician*, *43*, 253-260.
- ODEH, R. E., & EVANS, J. O. (1974). Algorithm AS70: The percentage points of the normal distribution. *Applied Statistics*, *23*, 96-97.
- ROTHSTEIN, H., BORENSTEIN, M., COHEN, J., & POLLACK, S. (1990). Statistical power analysis for multiple regression/correlation: A computer program. *Educational & Psychological Measurement*, *50*, 819-830.
- TIKU, M. L. (1967). Tables of the power of the F-test. *Journal of the American Statistical Association*, *62*, 525-539.
- TVERSKY, A., & KAHNEMAN, D. (1971). Belief in the law of small numbers. *Psychological Bulletin*, *76*, 105-110.
- ZELEN, M., & SEVERO, N. C. (1964). *Probability functions*. In M. Abramowitz & I. Stegun (Eds.), *Handbook of mathematical functions* (National Bureau of Standards, Applied Mathematics Series No. 55). Washington, DC: U.S. Government Printing Office.

### APPENDIX

#### Algorithm for Computation of Power

```

FUNCTION TO COMPUTE POWER
ADAPTED FROM COHEN (1988)

REAL FUNCTION: COMPUTEPOWER
D=ABS((MEAN1-MEAN2))/SDPRIME(SD1,SD2)
N=NPRIME(N1,N2)
DENOM=2*(N-1)+1.21*(ZALPHA-1.06)
NUM=D*(N-1)*SQR(2*N)
ZALPHA=ZFROMALPHA(ALPHA,TAILS)
Z=(NUM/DENOM)-ZALPHA
TEMP1=CUMZ(Z)
TEMP2=0
DO IF TAILS=2
NUM=-1*D*(N-1)*SQR(2*N)
Z=(NUM/DENOM)-ZALPHA
TEMP2=CUMZ(Z)
END DO
RESULT=TEMP1+TEMP2
END FUNCTION

```

## APPENDIX (Continued)

---

```

REAL FUNCTION: SDPRIME
RESULT=SQR((SD1*SD1)+(SD2*SD2))/2)
END FUNCTION

```

```

REAL FUNCTION: NPRIME
RESULT=(2*N1*N2)/(N1+N2)
END FUNCTION

```

```

FUNCTIONS TO COMPUTE AREA TO LEFT OF Z

```

```

REAL FUNCTION: CUMZ
IF Z<0 RESULT= PROBZ(Z,1)
RESULT=(.5-PROBZ(Z,1))+.5
END FUNCTION

```

```

REAL FUNCTION: PROBZ
CONSTANT: A1=.0705230784
CONSTANT: A2=.0422820123
CONSTANT: A3=.0092705272
CONSTANT: A4=.0001520143
CONSTANT: A5=.0002765672
CONSTANT: A6=.0000430638
Z=MIN(.7071067812*ABS(ZIN),10)
P1=.5*(1+Z*(A1+Z*(A2+Z*(A3+Z*(A4+Z*(A5+Z*(A6))))))^-16
P2=P1*2
IF TAILS=1 THEN RESULT=P1:EXIT
IF TAILS=2 THEN RESULT=P2:EXIT
END FUNCTION

```

```

FUNCTIONS TO COMPUTE Z FOR ALPHA

```

```

REAL FUNCTION: ZFROMALPHA
IF TAILS=1 AND ALPHA <=.5 THEN CUM=1-ALPHA
IF TAILS=1 AND ALPHA >.5 THEN CUM=.5
IF TAILS=2 THEN CUM=1-ALPHA/2
RESULT=ZFROMPROB(CUM,1)
END FUNCTION

```

```

REAL FUNCTION: ZFROMPROB
CONSTANT: A11=4.53642210148E-05
CONSTANT: A12= .0204231210245
CONSTANT: A13= .342242088547
CONSTANT: A14= .322232431088
CONSTANT: A15= .0038560700634
CONSTANT: A16= .10353775285
CONSTANT: A17= .531103462366
CONSTANT: A18= .588581570495
CONSTANT: A19= .099348462606
XP=P:IF TAILS=2 THEN XP=P/2
R=XP:IF XP>.5 THEN R=1-XP
IF R<1E-20 THEN Z=10: GOTO 70
Y=SQR(-2*LOG(R))
Z=Y-(((A11*Y+A12)*Y+A13)*Y+1)*Y+A14)/((((A15*Y+A16)*Y+A17)*Y+A18)*Y+A19)
IF XP<.5 THEN Z=-Z
RESULT=Z:EXIT
END FUNCTION

```

---

(Manuscript received November 7, 1991;  
revision accepted for publication June 3, 1992.)