Sampling Using Cohen's (1992) Two Tables

(1) Effect Sizes, d and r

d = the mean difference between two groups divided by the standard deviation for the data

Cohen's d = $\frac{Mean\ 1 - Mean\ 2}{Pooled\ SD}$

r = Pearson correlation between two variables

Examples of d

(a) Salary Differences

Females = $50,000
Males = $45,000
Standard deviation for salary = $10,000

Cohen's d = $\frac{Mean\ 1 - Mean\ 2}{Pooled\ SD}$ = $\frac{50{,}000 - 45{,}000}{10{,}000}$ = $\frac{5{,}000}{10{,}000}$ = .50

d = .50 which means females earn half a standard deviation more than males

(b) Salary Differences, again

Females = $50,000
Males = $58,000
Standard deviation for salary = $10,000

Cohen's d = $\frac{Mean\ 1 - Mean\ 2}{Pooled\ SD}$ = $\frac{50{,}000 - 58{,}000}{10{,}000}$ = $\frac{-8{,}000}{10{,}000}$ = -.80

d = -.80 which means females earn 8/10 a standard deviation less than males

(c) Salary Differences, once more

Females = $52,000
Males = $50,000
Standard deviation for salary = $10,000

Cohen's d = $\frac{Mean\ 1 - Mean\ 2}{Pooled\ SD}$ = $\frac{52{,}000 - 50{,}000}{10{,}000}$ = $\frac{2{,}000}{10{,}000}$ = .20

d = .20 which means females earn 2/10 a standard deviation more than males

Effect Size d, Magnitude According to Cohen
(See Table 1, p 157, of Cohen 1992)

.20 = small
.50 = medium
.80 = large

also, note that sign does not diminish effect size, so these are equivalent in strength

.20 = -.20 = small
.50 = -.50 = medium
.80 = -.80 = large

Examples of Pearson Correlation Coefficient, r

(a) Test Anxiety and Academic Self-efficacy

r = -.50

Negative correlation, as self-efficacy increases, test anxiety declines

(b) Student Ratings of Instructors and Level of Classroom Autonomy

r = .30

Positive correlation, as classroom autonomy increases, so too do mean instructor ratings by students

(c) Salary and Job Satisfaction

r = .10

Positive correlation, as classroom autonomy increases, so too do mean instructor ratings by students

Effect Size r, Magnitude According to Cohen
(See Table 1, p 157, of Cohen 1992)

.10 = small
.30 = medium
.50 = large

also, note that sign does not diminish effect size, so these are equivalent in strength

.10 = -.10 = small
.30 = -.30 = medium
.50 = -.50 = large

(2) Other Effect Sizes

Cohen (1992) Table 1, p. 157, lists a number of other effect sizes besides d and r, for example, f and f^2. Each of these can be viewed as equivalents to d and r for interpretation purposes.

Table 1
ES Indexes and Their Values for Small, Medium, and Large Effects

Test	ES index	Effect size Small	Medium	Large
1. m_A vs. m_B for independent means	$d = \frac{m_A - m_B}{\sigma}$	.20	.50	.80
2. Significance of product–moment r	r	.10	.30	.50
3. r_A vs. r_B for independent rs	$q = z_A - z_B$ where z = Fisher's z	.10	.30	.50
4. $P = .5$ and the sign test	$g = P - .50$	.05	.15	.25
5. P_A vs. P_B for independent proportions	$h = \phi_A - \phi_B$ where ϕ = arcsine transformation	.20	.50	.80
6. Chi-square for goodness of fit and contingency	$w = \sqrt{\sum_{i=1}^{k} \frac{(P_{1i} - P_{0i})^2}{P_{0i}}}$	.10	.30	.50
7. One-way analysis of variance	$f = \frac{\sigma_m}{\sigma}$	.10	.25	.40
8. Multiple and multiple partial correlation	$f^2 = \frac{R^2}{1 - R^2}$	.02	.15	.35

Note. ES = population effect size.

(3) Sample Size Table

Cohen (1992) provides a simple sample size figures Table 2, p. 158:

Table 2
N for Small, Medium, and Large ES at Power = .80 for α = .01, .05, and .10

	α								
	.01			.05			.10		
Test	Sm	Med	Lg	Sm	Med	Lg	Sm	Med	Lg
1. Mean dif	586	95	38	393	64	26	310	50	20
2. Sig *r*	1,163	125	41	783	85	28	617	68	22
3. *r* dif	2,339	263	96	1,573	177	66	1,240	140	52
4. *P* = .5	1,165	127	44	783	85	30	616	67	23
5. *P* dif	584	93	36	392	63	25	309	49	19
6. χ^2									
1*df*	1,168	130	38	785	87	26	618	69	25
2*df*	1,388	154	56	964	107	39	771	86	31
3*df*	1,546	172	62	1,090	121	44	880	98	35
4*df*	1,675	186	67	1,194	133	48	968	108	39
5*df*	1,787	199	71	1,293	143	51	1,045	116	42
6*df*	1,887	210	75	1,362	151	54	1,113	124	45
7. ANOVA									
2*g*[a]	586	95	38	393	64	26	310	50	20
3*g*[a]	464	76	30	322	52	21	258	41	17
4*g*[a]	388	63	25	274	45	18	221	36	15
5*g*[a]	336	55	22	240	39	16	193	32	13
6*g*[a]	299	49	20	215	35	14	174	28	12
7*g*[a]	271	44	18	195	32	13	159	26	11
8. Mult *R*									
2*k*[b]	698	97	45	481	67	30			
3*k*[b]	780	108	50	547	76	34			
4*k*[b]	841	118	55	599	84	38			
5*k*[b]	901	126	59	645	91	42			
6*k*[b]	953	134	63	686	97	45			
7*k*[b]	998	141	66	726	102	48			
8*k*[b]	1,039	147	69	757	107	50			

Note. ES = population effect size, Sm = small, Med = medium, Lg = large, diff = difference, ANOVA = analysis of variance. Tests numbered as in Table 1.
[a] Number of groups. [b] Number of independent variables.

Power = .80

This means a statistical test run with the sample sizes provided and with the effect size assumed (small, medium, or large) has an 80% chance of finding that effect.

α = .01, .05, or .10

Alpha (α) is the probability of making a Type 1 error in hypothesis testing (incorrectly concluding that you found an effect, difference, or relationship when there really is not effect, difference, or relationship in the population).

A value of .01 means this error is expected to occur 1 out of 100 tests.
A value of .05 means this error is expected to occur 5 out of 100 tests.
A value of .10 means this error is expected to occur 10 out of 100 tests.

Most researchers use .05 unless they have large samples and then .01 is used.

1. Mean dif

Per-group sample size needed for two-group t-test or one-way ANOVA with two groups.

This means the total sample size will be twice the number presented in the table.

For example, alpha = .05 and large effect size (d) corresponds to a sample size of 26 for each group, thus total sample of 26 + 26 = 52 for both groups combined.

2. Sig r

Sample size needed for testing Pearson correlation.

For example, if we wanted to find a medium correlation of r = .30, with alpha = .01, we would need a sample of 125

7. ANOVA

This section provides sample sizes per group in ANOVA.

The numbers below the title ANOVA represent the number of groups, 2g = two groups, 3g = three groups, 4g = four groups, etc.

As with Mean Dif above, the sample sizes provided are per group, so to find the total sample size one must multiple the sample size provided by the number of groups.

For example, alpha = .05, effect size = medium, and there are 3 groups, the sample size reported is 52, so the total sample size would be 52 x 3 = 156.

8. Mult R

This section provides sample sizes for multiple regression. Unfortunately, this section is not very helpful because the null hypothesis Cohen used to determine sample size in regression is for the overall model fit:

Ho: $R^2 = 0.00$

In words: Does the collective regression model (i.e., all predictors variables included) predict more variance in the dependent variable than would be expected by chance?

A more useful sample size for regression would be based upon individual predictors, that is, determining sample size needed to assess small, medium, or large effects for one of the predictors in the regression equation:

Ho: $\beta_i = 0.00$

In words: Does the predictor i contribute the the regression model (i.e., reduce error in predicting the dependent variable) over and above the contributions made by other predictors in the regression model?

The problem with sample sizes for the overall model is that these samples are too small to confidently test for the significance of individual predictors. Thus, hypothesis testing to learn which predictors are related to the dependent variable will be underpowered if using the values provided in Cohen's table for regression.

Nevertheless, if one wished to test for overall model fit, rather than for individual predictors, the sample sizes provided by Cohen apply.

The numbers below the title Mult R indicates the number of variables presented in the analysis (this includes dummy variables two for qualitative predictors).

The sample sizes provided represent the total sample size, so no need to multiply tabled sample sizes to obtain the total sample size.

For example, if we run a regression with four predictors with alpha = .01, effect size at medium, a total sample of n = 118 would be needed.

(4) Examples

What sample size would be needed for each of the following?

(a) Is there a difference in spelling test scores between boys and girls in Mrs. Platt's classroom?

Effect size d = .50

alpha = .01

(b) Is there a relationship between SAT scores and and freshman year GPA?

r = .30

alpha = .01

(c) Are there differences in standardized US history scores among students in Mrs. Jones', Mr. Scarff's, and Mrs. McKenna's classes?

Effect size = small

alpha = .01

(d) Are there differences in standardized US history scores among students in Mrs. Jones', Mr. Scarff's, and Mrs. McKenna's classes?

Effect size = small

alpha = .05

(e) Are there differences in standardized US history scores among students in Mrs. Jones', Mr. Scarff's, and Mrs. McKenna's classes?

Effect size = medium

alpha = .05

(f) Are there differences in standardized US history scores among students in Mrs. Jones', Mr. Scarff's, Mrs. McKenna's and Mr. Lattner's classes?

Effect size = medium

alpha = .05

(g) Do females in single-sex classes perform better than females in coeducational classes in high school biology? Below are end of course test scores in 9^{th} grade biology for two classes, one taught with female-only and one with both males and females present.

Effect size = small

alpha = .05

(h) Which is the better predictor of one's job satisfaction, sense of autonomy, relatedness, or competence on the job?

Effect size = small

alpha = .01

(i) Do teacher-student ratio and average teacher salary simultaneously predict student performance on the mathematics section of the SAT? Which of these variables is the better predictor, if either, once the other is controlled or taken into account?

Effect size = large

alpha = .05

(5) Example Answers

(a) Is there a difference in spelling test scores between boys and girls in Mrs. Platt's classroom?

Effect size d = .50

alpha = .01

n = 95 + 95 = 190

(b) Is there a relationship between SAT scores and and freshman year GPA?

r = .30

alpha = .01

n = 125

(c) Are there differences in standardized US history scores among students in Mrs. Jones', Mr. Scarff's, and Mrs. McKenna's classes?

Effect size = small

alpha = .01

n = 464 x 3 = 1392

(d) Are there differences in standardized US history scores among students in Mrs. Jones', Mr. Scarff's, and Mrs. McKenna's classes?

Effect size = small

alpha = .05

n = 322 x 3 = 966

(e) Are there differences in standardized US history scores among students in Mrs. Jones', Mr. Scarff's, and Mrs. McKenna's classes?

Effect size = medium

alpha = .05

n = 52 x 3 = 156

(f) Are there differences in standardized US history scores among students in Mrs. Jones', Mr. Scarff's, Mrs. McKenna's and Mr. Lattner's classes?

Effect size = medium

alpha = .05

n = 45 x 4 = 180

(g) Do females in single-sex classes perform better than females in co-educational classes in high school biology? Below are end of course test scores in 9^{th} grade biology for two classes, one taught with female-only and one with both males and females present.

Effect size = small

alpha = .05

n = 393 x 2 = 786

(h) Which is the better predictor of one's job satisfaction, sense of autonomy, relatedness, or competence on the job?

Effect size = small

alpha = .01

n = 780

(i) Do teacher-student ratio and average teacher salary simultaneously predict student performance on the mathematics section of the SAT? Which of these variables is the better predictor, if either, once the other is controlled or taken into account?

Effect size = large
alpha = .05

n = 30