DOCUMENT RESUME

ED 336 411 TM 017 177

| | |
|---|---|
| AUTHOR | Sawilowsky, Shlomo S.; Hillman, Stephen B. |
| TITLE | Sample Size Tables, "t" Test, and a Prevalent Psychometric Distribution. |
| INSTITUTION | Wayne State Univ., Detroit, Mich. Coll. of Education. |
| PUB DATE | Aug 91 |
| NOTE | 9p.; Paper presented at the Annual Meeting of the American Psychological Association (99th, San Francisco, CA, August 16-20, 1991). |
| PUB TYPE | Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150) |
| EDRS PRICE | MF01/PC01 Plus Postage. |
| DESCRIPTORS | Mathematical Models; Monte Carlo Methods; *Psychological Studies; *Psychometrics; Research Methodology; *Sample Size; *Statistical Distributions; Surveys; Tables (Data) |
| IDENTIFIERS | *Power (Statistics); *T Test; Type I Errors; Type II Errors |

ABSTRACT

Psychology studies often have low statistical power. Sample size tables, as given by J. Cohen (1988), may be used to increase power, but they are based on Monte Carlo studies of relatively "tame" mathematical distributions, as compared to psychology data sets. In this study, Monte Carlo methods were used to investigate Type I and Type II error properties of the independent samples "t" test under a "discrete mass at zero with gap" data set to determine if the sample size tables given by Cohen yield correct results. Monte Carlo methods were used with a FORTRAN program to sample with replacement from a population of 516 responses to a survey regarding the age at which subjects first used cigarettes. Ten sample sizes were randomly drawn: (1) n1=5, n2=15; (2) n1=10, n2=10; (3) n1=10, n2=30; (4) n1=20, n2=20; (5) n1=15, n2=45; (6) n1=30, n2=30; (7) n1=20, n2=60; (8) n1=40, n2=40; (9) n1=30, n2=90; and (10) n1=60, n2=60. For the smallest unbalanced sample size (5,15), the "t" test was generally not robust. For the remaining sample sizes, results were in agreement with normal curve theory. When confronted with non-normal data sets, psychology researchers do not need to make any modifications to Cohen's (1988) tables when making sample size determinations. Two graphs illustrate the study. A 12-item list of references is included. (SLD)

Wayne State University

College of Education
Detroit, Michigan 48202

Division of Theoretical and
Behavioral Foundations

Educational and Clinical Psychology

Laboratory for Research on Adolescence

ED336411

# SAMPLE SIZE TABLES, *t* TEST, AND A PREVALENT PSYCHOMETRIC DISTRIBUTION

SHLOMO S. SAWILOWSKY
Educational Evaluation & Research

STEPHEN B. HILLMAN
Educational & Clinical Psychology

Laboratory For Research On Adolescence
College of Education
Wayne State University



Paper presented at the 99th annual meeting of the American Psychological Association, Division 5, Evaluation, Measurement, and Statistics, San Francisco, California, August, 1991

TM017177

SAMPLE SIZE TABLES, $t$ TEST, AND
A PREVALENT PSYCHOMETRIC DISTRIBUTION

SHLOMO S. SAWILOWSKY STEPHEN B. HILLMAN

Laboratory For Research On Adolescence
College of Education
Wayne State University

## ABSTRACT

Psychology studies often have low statistical power. Sample size tables (Cohen, 1988) may be used to increase power, but they are based on Monte Carlo studies of relatively "tame" mathematical distributions, as compared to psychology data sets. A prevalent psychometric measure distribution, the "discrete mass at zero with gap", occurs with "first use" variables. The Type II error properties of the independent samples $t$ test on real psychometric data of this type were in agreement with normal curve theory. Thus, in making sample size determinations, psychology researchers do not need to make any modifications for this common distribution.

# SAMPLE SIZE TABLES, *t* TEST, AND A PREVALENT PSYCHOMETRIC DISTRIBUTION

In an analysis of 85 published psychotherapy outcome studies from 1984-1986, Kazdin and Bass (1989) found the power to detect differences between two or more treatments was weak. Similarly, Rossi (1990) calculated power for 6,155 statistical tests performed in psychological research published in 1982 and concluded the power was very low. A simple method to maximize statistical power is through sample size determinations, such as through tables given by Cohen (1988).

A serious question has arisen, however, that may directly effect sample size calculations. The question pertains to the robustness of parametric statistics such as the *t* and F tests to departures from population normality. Monte Carlo studies by Boneau (1960, 1962) and Glass, Peckham and Sanders (1972), and the text by Scheffe (1959) are often cited as evidence of the robustness of these tests. Bradley (1968, 1977, 1978, 1982) countered that their simulation studies were limited to the investigation of mathematically convenient distributions that are relatively "tame" in comparison to distributions obtained in psychology.

Micceri (1986, 1989) obtained 440 large sample data sets from research published in 1982-1984 that dramatically underscores Bradley's concerns. One hundred and twenty-five of the 440 large sample data sets were obtained from psychometric measures, including Minnesota Multiphasic Personality Inventory scales, Mallory test of visual hallucinations, and a variety of measures of anger, anxiety, curiosity, locus of control, masculinity/femininity, satisfaction, sociability, etc. A mere 3% of these data sets were considered near Gaussian. Micceri (1989) concluded that studies by Boneau and others of the robustness of the *t* and F tests under smooth mathematical curves were inconclusive, because "none of these comparisons occurs in real life" (p. 164).

Figure 1 is a histogram depicting a prevalent psychometric measure distribution described by Micceri as "discrete mass at zero with gap". This example represents the responses of 516 adolescents to a survey question asking their age when they first began to smoke cigarettes. Distributions such as these typically occur with "first use" or "onset" variables in psychological research. Researchers may only be interested in responses above zero. Consider the two paradigms emerging in recent substance use literature: "User/Abuser" and "Nonuser/User/Abuser". For studies such as Long and Scherl (1984) and Hillman and Sawilowsky (1991), the percent of nonusers would be reported, but hypotheses (such as comparing means) would be tested only on user and abuser categories. Although this part of the curve is a bit flatter than the normal curve, little doubt is raised regarding Type II error properties. For the latter case, however, such as Shedler and Block (1990) who studied "Abstainers, Experimenters, and Frequent Users", it would be appropriate to compare means for all three groups even though the nonuser scores are discrete zeros. Figure 1 probably depicts a population shape unimagined by psychology researchers who think in terms of normal curve theory.

## PURPOSE OF THE STUDY

The issue raised by Micceri applies equally to Type II error (or power). Cohen (1988) relied primarily on Scheffe (1959) and Boneau (1960, 1962) in preparing sample size tables. Because these studies used mathematically convenient and tame distributions, the issues raised by Bradley (1968) and Micceri (1989) bring into question the validity of the tables. The purpose of this study, then, is to use Monte Carlo methods to investigate the Type I and Type II error properties of the independent samples *t* test under a "discrete mass at zero with gap" data set to determine if sample size tables in Cohen (1988) yield correct results.

## METHODOLOGY

Monte Carlo methods were used with a FORTRAN program to sample with replacement from a population of 516 responses to a survey question regarding the age of first use of cigarettes. Scores were also sampled from a Gaussian distribution to demonstrate the adequacy of the simulation. Random responses were drawn for sample sizes $(n_1, n_2)$ = (5,15), (10,10), (10,30), (20,20), (15,45), (30,30), (20,60), (40,40), (30,90), and (60,60).

The Type I error rates were obtained by computing the *t* statistic on each sample pair with ten thousand (10,000) repetitions performed for the .10, .05, and .01 alpha levels. The robustness of the independent samples *t* test with respect to Type II error was investigated as follows: Let $X_{1i}$ and $X_{2j}$ be observations in two random samples with mean $\mu_1$ and standard deviation $\sigma_1$. Transformed variables called $X'_{1i}$ and $X'_{2j}$ were generated by

$$(1) \qquad X'_{1i} = X_{1i} - \mu_1 \qquad\qquad i = 1, \ldots, n_1$$

$$(2) \qquad X'_{2j} = c\,(X_{2j} - \mu_1) + k\,\sigma_1 \qquad\qquad j = 1, \ldots, n_2$$

where c and k are constants.

Hypotheses of shift in location parameters were investigated by making c = 1 and k equal to a constant of $.2\sigma$, $.5\sigma$, $.8\sigma$, and $1.2\sigma$, where $\sigma$ represents the standard deviation of the distribution, to $X_{2j}$: Hypotheses of shift in mean plus increases in variance were investigated by using the same values of k, but c was made equal to $\sqrt{2}$, $\sqrt{3}$, and 2. Unbalanced layouts were not investigated for $c \neq 1$, as Cohen (1988) noted that for both unequal variances and unequal sample sizes the tabled power values "may be greatly in error" (p. 44).

## RESULTS AND DISCUSSION

For the smallest unbalanced sample size (5,15) the *t* test was generally not robust. With nominal alpha at .10, the lower tail rejected only .023 and the upper tail was .050; with alpha at .05, the lower tail rejected only .009, while the upper tail was slightly liberal at .029; and at the .01 alpha level the lower tail was conservative at .001 and the upper tail was liberal at .007. However, for the remaining sample sizes, results were in agreement with normal curve theory. As noted in Figure 2, the power curves for the nonnormal data set for sample size (5,15) indicated a slight power loss of.01. At the larger sample sizes the power was virtually identical to that expected under normal curve theory. Thus, when confronted with nonnormal data sets such as this, psychology researchers need not make any modifications to Cohen's (1988) tables when making sample size determinations.

## REFERENCES


Boneau, C. A. (1960). The effects of violations of assumptions underlying the t test. Psychological Bulletin, 57, 49-64.

Bradley, J. V. (1968). Distribution-free statistical tests. Englewood-Cliffs, N.J.: Prentice Hall.

Bradley, J. V. (1977). A common situation conducive to bizzare distribution shape. The American Statistician, 31, 147-150.

Bradley, J. V. (1978). Robustness? British Journal of Mathematical and Statistical Psychology, 31, 144-152.

Bradley, J. V. (1982). The insidious L-shaped distribution. Bulletin of the Psychometrics Society, 20, 85-88.

Cohen, J. Statistical power analysis for the behavioral sciences (2nd ed). Hillsdale, N.J.: Lawrence Erlbaum.

Hillman, S. B., & Sawilowsky, S. S. (1991). Multidimensional differences between adolescent substance abusers and users. Psychological Reports, 68, 115-122.

Kazdin, A. E. & Bass, D. (1989). Power to detect differences between alternative treatments in comparative psychotherapy outcome research. Journal of Consulting and Clinical Psychology, 57, 138-147.

Micceri, T. (1986, November). A futile search for that statistical chimera of normality. Florida Educational Research Association, Tampa, Fl.

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

Rossi, J. S. (1990). Statistical power of psychological research: what have we gained in 20 years? Journal of Consulting and Clinical Psychology, 58, 646-656.

Scheffe, H. (1959). The analysis of variance. N.Y.: Wiley.

Figure 1.
*Age of First Use of Cigarettes*

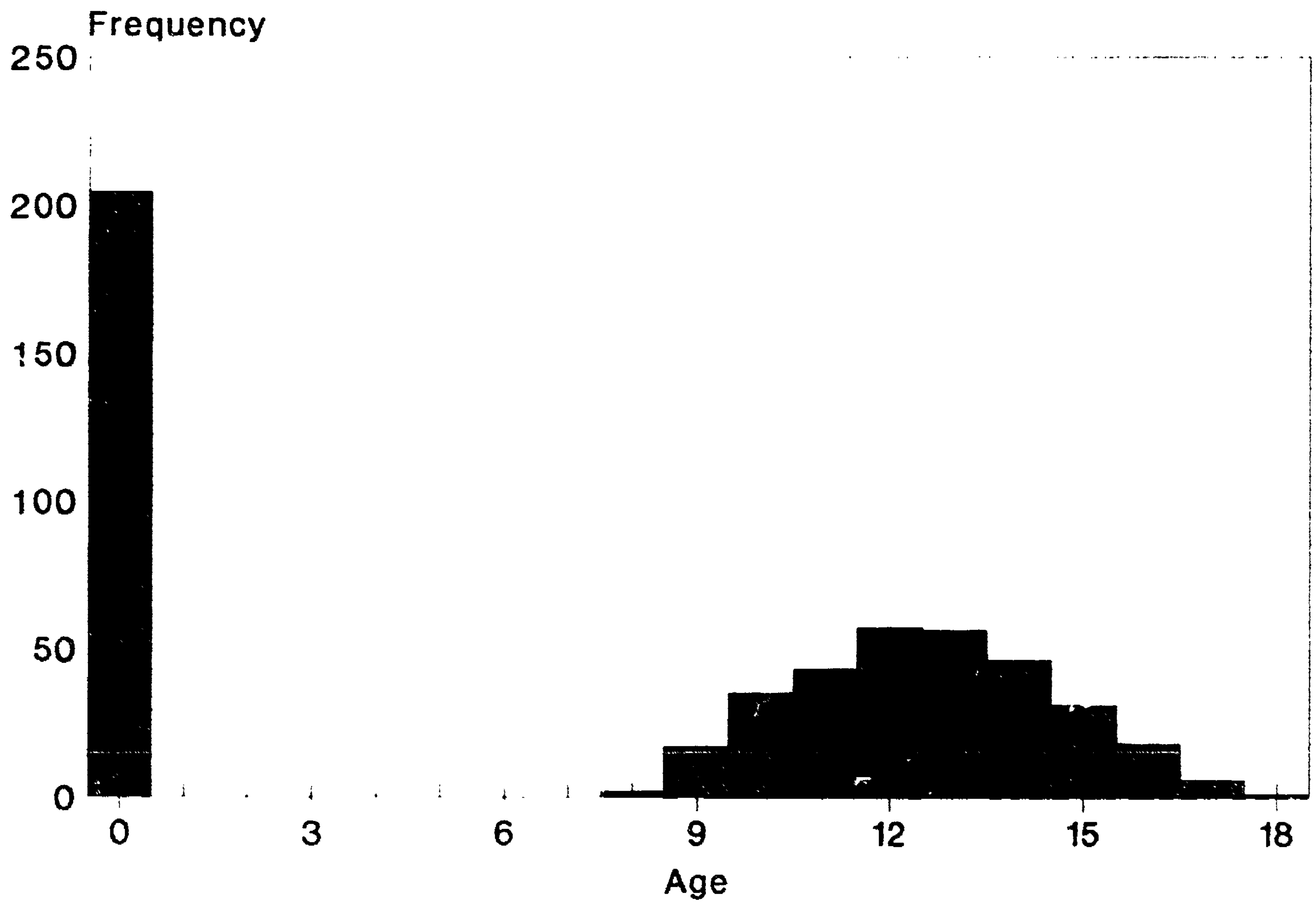

FIGURE 2

*One-Tail t-Test Power Curves For n1,n2 = (5,15); Alpha = .05*