Table 1. Required Sample Sizes for $\alpha = .05$

	ρ			
δ	.2	.3	.5	.8
.05	38,400	16,200	4,800	900
.10	9,600	4,050	1,200	225
.20	2,400	1,000	300	57

Next, define

$$T = [\hat{\beta} - \beta]/[\sigma_y^2 (1 - \rho^2)/(n - 1)\, \sigma_x^2]^{1/2}. \quad (1.4)$$

Then Thigpen and Paulson (1974) gave a proof (in a more general setting) that T has Student's t distribution with $n - 1$ df. The t statistic in (1.4) may be used to find the necessary sample size for estimating β from $\hat{\beta}$ so that the relative error will be controlled with predetermined probability.

2. THE SAMPLE-SIZE PROBLEM

Suppose that we are to sample from a bivariate normal population and wish to estimate the regression slope, β, so that the relative error of estimation

$$|(\hat{\beta} - \beta)/\beta| \quad (2.1)$$

will be less than δ, say, with probability $1 - \alpha$. With proper substitutions, the t statistic in (1.4) may be rewritten as

$$T = [(\hat{\beta} - \beta)/\beta]\,[(n - 1)\rho^2/(1 - \rho^2)]^{1/2}. \quad (2.2)$$

Letting $t_{\alpha/2}$ be such that $\Pr(T > t_{\alpha/2}) = \alpha/2$ for Student's t distribution with $n - 1$ df, it is seen that (2.1) exceeds $t_{\alpha/2}\,[(1 - \rho^2)/(n - 1)\rho^2]^{1/2}$ with probability α.

Now, if we desire that, for preselected δ and α,

$$\Pr[(\hat{\beta} - \beta)/\beta < \delta] = 1 - \alpha, \quad (2.3)$$

we shall need

$$t_{\alpha/2}[(1 - \rho^2)/(n - 1)\rho^2]^{1/2} = \delta. \quad (2.4)$$

Solving (2.4) for n we find that

$$(n - 1)^{1/2} = t_{\alpha/2}[(1 - \rho^2)/\rho^2]^{1/2}/\delta. \quad (2.5)$$

3. AN EXAMPLE

Suppose that we wish $1 - \alpha$ to equal .95. Then if we take $t_{\alpha/2} = 2$ as an approximation in (2.5) we can solve for the required sample size for given ρ so that $\hat{\beta}$ will have no more relative error than δ with probability .95. Some easy calculations give rise to Table 1. For small values of ρ, the required sample sizes are remarkably large. This is because as ρ approaches 0, β approaches 0, in which case small percentage errors are going to be difficult to achieve.

[*Received December 1985.*]

REFERENCE

Thigpen, C. C., and Paulson, A. S. (1974), "A Multiple Range Test for Analysis of Covariance," *Biometrika*, 61, 479–484.

A Note on Confidence Intervals for Proportions in Finite Populations

JOHN P. BUONACCORSI*

An exact confidence interval for the number or proportion of successes in a finite population is developed using the standard technique of inverting a family of tests. The resulting procedure is compared with two methods available in the sampling literature and is shown to be equivalent to one of the methods and superior to the other method.

KEY WORDS: Family of tests; Hypergeometric distribution; Sampling.

1. INTRODUCTION

A common problem is that of estimating the number or proportion of units in a finite population that possess a particular characteristic of interest. Classical examples include the number of persons holding a particular opinion or the number of defective items in a lot of items.

Suppose that the population is of size N, and let A_0 denote the number of units having the characteristic of interest and let $P = A_0/N$. This article is concerned with the construction of exact confidence intervals for A_0 and/or P. For convenience, I provide results only for A_0, since intervals for P are obtained by simply dividing by N.

Assume that a simple random sample without replacement of size n is selected. Let X = number of units in the sample having the characteristic of interest. For "large" n, an approximation to the distribution of X is used to generate approximate confidence intervals. Exact confidence intervals are based on the exact distribution of X, which is hypergeometric with parameters N, A_0, and n. These intervals are exact in the following sense. If we let $C(x)$ denote a random interval and write $\Pr_A$ for probabilities under $A_0 =$

*John P. Buonaccorsi is Assistant Professor, Department of Mathematics and Statistics, University of Massachusetts, Amherst, Massachusetts 01003. He wishes to thank Bill Link for some helpful comments.

A, then $\Pr_A[A \in C(X)] \geq 1 - \alpha$ for $A = 0, 1, \ldots, N$ $(0 < \alpha < 1)$.

Write $C(x) = [L(x), U(x)]$ to denote the observed confidence interval, where x denotes the realization of the random variable X. In some places where there is no danger of confusion, I write simply L and U.

Section 2 discusses the available methods, and Section 3 develops an exact interval by inverting a family of tests. Section 4 shows the equivalence of the method in Section 3 with one of the available methods. These are in turn shown to be superior to an often-used technique. Finally, an example is presented.

In what follows, I partition the value α up equally between each of the tails. This is not essential, similar results being available for α_1, α_2, both in $(0, .5)$, such that $\alpha_1 + \alpha_2 = \alpha$.

2. AVAILABLE METHODS

A perusal of the sampling literature yields the following methods for finding exact confidence intervals for A_0.

Cochran (1976, p. 57) used

$$L(x) = \text{largest integer } A \text{ such that } \Pr_A[X \geq x] \leq \alpha/2 \tag{2.1}$$

and

$$U(x) = \text{smallest integer } A \text{ such that } \Pr_A[X \leq x] \geq \alpha/2. \tag{2.2}$$

When $x = 0$, no A satisfies (2.1) and $L(x)$ is set equal to 0. Similarly, when $x = n$, $U(x)$ is set equal to N.

Sukhatme, Sukhatme, Sukhatme, and Asok (1984, p. 46) suggested choosing L and U to satisfy $\Pr_L[X \geq x] = \alpha/2$ and $\Pr_U[X \leq x] = \alpha/2$, respectively. Such solutions do not necessarily exist, however, although if N is large, one can get very close to finding such solutions. This is essentially what was done by Chung and Delury (1950) in constructing charts for some large N; the smallest N in these charts is 500.

Konijn (1973, p. 79) described a method that takes $L(x)$ to be the largest integer A such that for all $K < A$, $\Pr_K[X \geq x] \leq \alpha/2$ and $U(x)$ is the smallest integer A such that for all $K > A$, $\Pr_K[X \leq x] \leq \alpha/2$. This is readily seen to be equivalent to

$$L(x) = \text{smallest } A \text{ such that } \Pr_A[X \geq x] > \alpha/2 \tag{2.3}$$

and

$$U(x) = \text{largest } A \text{ such that } \Pr_A[X \leq x] > \alpha/2. \tag{2.4}$$

This method appears also to be the exact method referred to by Katz (1953).

The International Mathematical and Statistical Libraries (1984) subroutine SSPAND uses Equations (2.1) and (2.2) but with a Poisson, binomial, or normal approximation to the distribution of X.

3. INVERTING TESTS

A common technique for generating a confidence interval is to invert a family of hypothesis tests for the parameter of interest (see Bickel and Doksum 1977, p. 177). A family of level-α tests for $H_0{:}A_0 = A$ versus $H_A{:}A_0 \neq A$ is as follows. Let

$$K_1(A) = \text{smallest integer } j \text{ such that } \Pr_A[X \geq j] \leq \alpha/2$$

and

$$K_2(A) = \text{largest integer } j \text{ such that } \Pr_A[X \leq j] \leq \alpha/2.$$

Note that both K_1 and K_2 are nondecreasing functions of A and $K_1 > K_2$. Further, as shown in the Appendix, as we move from A to $A + 1$, each of K_1 and K_2 changes by at most 1. Finally, note that $K_1(A) - K_2(A) \geq 2$.

If we adopt the decision rule to reject H_0 if $X \leq K_2(A)$ or $X \geq K_1(A)$, then for each A the resulting test is of level α; that is, under H_0 Pr(reject H_0) $\geq \alpha$. A $100(1 - \alpha)\%$ confidence interval for A_0 is given by

$$C(x) = \{A \text{ such that } H_0{:}A = A_0 \text{ is accepted}\}$$

or

$$C(x) = \{A \text{ such that } K_2(A) < x < K_1(A)\}.$$

It is straightforward to show (see the example in Sec. 3 for illustration) that $C(x)$ is given by

$$L(x) = \text{smallest } A \text{ such that } K_1(A) = x + 1$$

and

$$U(x) = \text{largest } A \text{ such that } K_2(A) = x - 1. \tag{3.1}$$

The equalities on the right result from the previously noted fact that jumps in K_1 and K_2 are at most 1.

4. COMPARISON OF THE METHODS

We refer to the different methods as the C (Cochran), K (Konijn), and T (test) method, respectively. It will be con-

Table 1. Cumulative Probabilities $K_1(A)$ and $K_2(A)$ for $N = 10$, $n = 4$, $\alpha = .10$

A	0	1	2	3	4	5	6	7	8	9	10
$\Pr_A[X \leq 0]$	1	.6	.33	.167	.071	.024	.005	0	0	0	0
$\Pr_A[X \leq 1]$	1	1	.867	.667	.452	.262	.119	.033	0	0	0
$\Pr_A[X \leq 2]$	1	1	1	.967	.881	.738	.548	.333	.133	0	0
$\Pr_A[X \leq 3]$	1	1	1	1	.995	.976	.929	.833	.667	.4	0
$\Pr_A[X \leq 4]$	1	1	1	1	1	1	1	1	1	1	1
$K_2(A)$	−1	−1	−1	−1	−1	0	0	1	1	2	3
$K_1(A)$	1	2	3	3	4	4	5	5	5	5	5

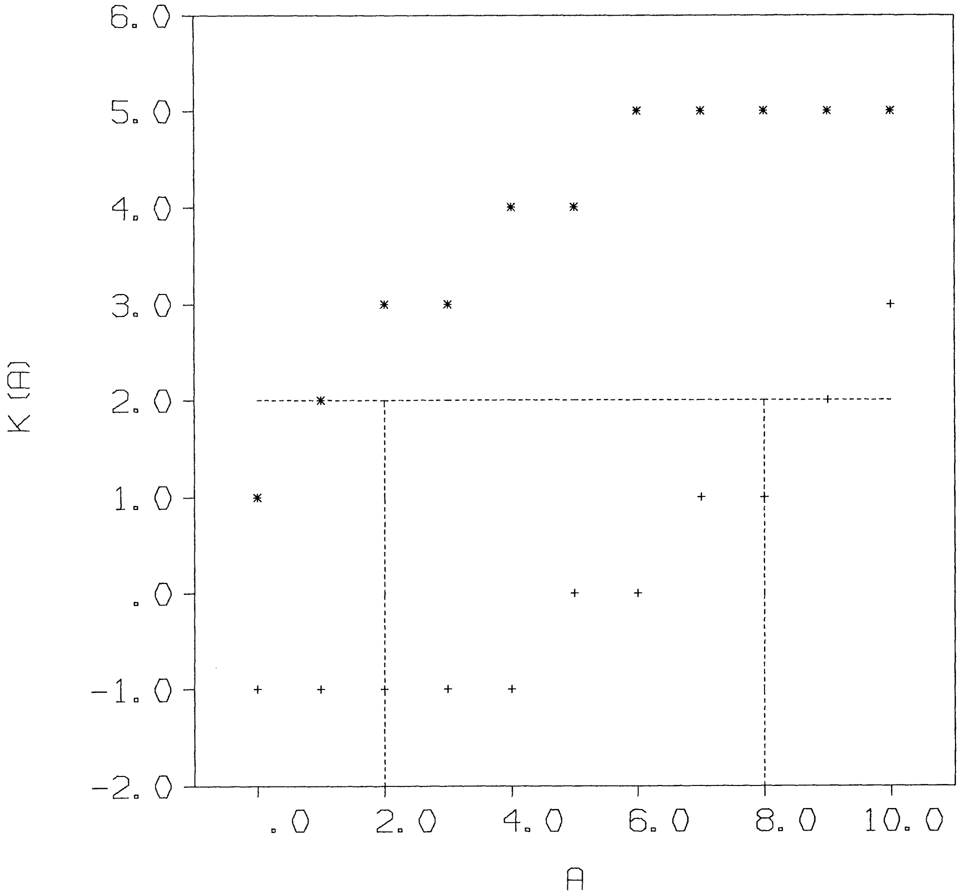

*Figure 1. Cutoff Points Used for the T Method and Construction of the T-Method Interval for $x = 2$: +, $K_2(A)$; *, $K_1(A)$.*

venient to denote the respective intervals by $[LC, UC]$, $[LK, UK]$, and $[LT, UT]$.

We have the following results.

1. The K method and the T method are equivalent. We show the equivalence of $UT(x)$ and $UK(x)$, a similar argument applying to the lower bounds. By (3.1), $K_2(UT) = x - 1$. This means that $\Pr_{UT}[X \le x - 1] \le \alpha/2$ and $\Pr_{UT}[X \le x] > \alpha/2$. We also have that $\Pr_{UT+1}[X \le x - 1] < \alpha/2$. If $\Pr_{UT+1}[X \le x] \ge \alpha/2$, then UT no longer satisfies (3.1). Hence $\Pr_{UT+1}[X \le x] \le \alpha/2$, which, since $\Pr_{UT}[X \le x] > \alpha/2$, implies by (2.4) that $UT = UK$.
2. The interval produced by the T (or K) method is a proper subset of the interval arising from the C method. More specifically, $LT(x) = LC(x) + 1$, for $x > 0$ with $LT(0) = LC(0) = 0$. Moreover, $UT(x) = UC(x) - 1$, for $x < n$ with $UT(n) = UC(n) = N$.

These results follow immediately from inspection of Equations (2.1), (2.2), (2.3), and (2.4).

5. EXAMPLE

To illustrate, consider the case with $N = 10$, $n = 4$, and $\alpha = .10$. Table 1 contains the cdf of X, $K_1(A)$, and $K_2(A)$ for each value of A. The values of K_1 and K_2 are also presented graphically in Figure 1. Table 2 contains the confidence intervals for each of the methods. The construction of the T method for $x = 2$ is represented in Figure 1.

6. DISCUSSION

The exact confidence interval obtained by inverting a family of tests was shown to be equivalent to the method used by Konijn (1973) and superior to the method used by Cochran (1976). Using one-tailed tests for A_0, upper and lower confidence bounds can be derived. These will be uniformly most accurate (Bickel and Doksum 1977, p. 206), since the corresponding tests are uniformly most powerful (see Lehmann 1959, p. 70).

Computationally, for moderate N, the interval $[LT, UT]$ is not difficult to obtain if one has available a routine that will compute hypergeometric probabilities. A reasonable approach would be to use as initial values the confidence interval for A_0 obtained with a normal approximation. For many situations the tables of Lieberman and Owen (1961) can be used.

Table 2. Observed Confidence Intervals for the Example

	K or T method		C method	
x	LT(x)	UT(x)	LC(x)	UC(x)
0	0	4	0	5
1	1	6	0	7
2	2	8	1	9
3	4	9	3	10
4	6	10	5	10

APPENDIX: JUMPS IN K_1 AND K_2

Here I show that each of $K_1(A + 1) - K_1(A)$ and $K_2(A + 1) - K_2(A)$ is at most 1. I consider K_2 a similar argument applying to K_1.

By definition $\Pr_A[X \le K_2(A)] \le \alpha/2$ and $\Pr_A[X \le K_2(A) + 1] > \alpha/2$. For $K_2(A + 1) - K_2(A) \ge 2$, we need

$$\Pr_{A+1}[X \le K_2(A) + 2] \le \alpha/2 < \Pr_A[X \le K_2(A) + 1]. \quad (A.1)$$

It is sufficient then to show that for all nonnegative integers j, $\Pr_A[X \le j] \le \Pr_{A+1}[X \le j + 1]$, which implies that (A.1) cannot hold.

Consider an urn with A red balls, 1 green ball, and $N - A - 1$ blue balls. Sample n balls at random without replacement, and let W = number of red balls observed and Y = number of green balls observed. Then W is hypergeometric (N, A, n) and $Z = W + Y$ is hypergeometric $(N, A + 1, n)$. We have

$$\Pr[Z \le j + 1] = \Pr[W \le j] + \Pr[W = j + 1, Y = 0] \ge \Pr[W \le j].$$

But in the notation of the article, $\Pr[Z \le j + 1] = \Pr_{A+1}[X \le j + 1]$ and $\Pr[W \le j] = \Pr_A[X \le j]$. This proves the result.

[*Received December 1985. Revised February 1987.*]

REFERENCES

Bickel, P. J., and Doksum, K. A. (1977), *Mathematical Statistics: Basic Ideas and Selected Topics*, San Francisco: Holden-Day.

Chung, H. H., and Delury, D. B. (1950), *Confidence Limits for the Hypergeometric Distribution*, Toronto, Ontario: University of Toronto Press.

Cochran, W. G. (1976), *Sampling Techniques* (3rd ed.), New York: John Wiley.

Katz, L. (1953), "Confidence Intervals for the Number Showing a Certain Characteristic in a Population When Sampling Is Without Replacement," *Journal of the American Statistical Association*, 48, 256–261.

Konijn, H. S. (1973), *Statistical Theory of Sample Survey Design and Analysis*, Amsterdam: North-Holland.

Lehmann, E. L. (1959), *Testing Statistical Hypotheses*, New York: John Wiley.

Lieberman, G. J., and Owen, D. B. (1961), *Tables of the Hypergeometric Probability Distribution*, Stanford, CA: Stanford University Press.

Sukhatme, P. V., Sukhatme, B. V., Sukhatme, S., and Asok, C. (1984), *Sampling Theory of Surveys With Applications*, Ames: Iowa State University Press & Indian Society of Agricultural Statistics.

Characterization of Risk Sets for Simple Versus Simple Hypothesis Testing

DANIEL Q. NAIMAN*

Consider testing a simple hypothesis H_0: $F = F_0$ against a simple alternative H_1: $F = F_1$ when a random variable X having distribution function F is observed. The risk region is defined to be the set of ordered pairs of error probabilities $(E_0[\phi(X)], E_1[1 - \phi(X)])$ as ϕ ranges throughout all tests. This set plays an important role in the Neyman–Pearson lemma, which characterizes the class of most powerful tests of H_0 versus H_1. In a typical introductory course in hypothesis testing, the risk region is shown to be a closed convex subset of the unit square $[0, 1] \times [0, 1]$ that is symmetric about (½, ½) and contains the points (1, 0) and (0, 1). In this article, it is shown that any set with these properties is the risk region for some pair of hypotheses. In fact, one can take F_0 to be the uniform distribution on [0, 1]. This result provides an exercise that illustrates the applicability of some results about convexity and absolute continuity that would be presented in a real analysis course at the level of Royden (1968).

KEY WORDS: Hypothesis test; Risk region; Neyman–Pearson lemma.

1. INTRODUCTION

The most basic form of hypothesis testing problem is one in which a simple null hypothesis H_0: $F = F_0$ is tested against a simple alternative hypothesis H_1: $F = F_1$, where F_0, F_1 are given distribution functions and a random variable having distribution F is observed.

A test of H_0 versus H_1 is a random variable of the form $\phi(X)$ taking values in [0, 1], for which one interprets $\phi(x)$ as the probability of rejecting H_0 when $X = x$ has been observed. To evaluate the performance of a given test, it is customary to consider the Type I error probability $E_0[\phi(X)]$, the probability of a false rejection, and the Type II error probability $E_1[1 - \phi(X)]$, the probability of a false acceptance. The risk region, which is defined to be the set of all ordered pairs of error probabilities $(E_0[\phi(X)], E_1[1 - \phi(X)])$ as ϕ ranges throughout all tests, plays an important role in the Neyman–Pearson lemma (see Ferguson 1967; Lehmann 1986), which characterizes the class of most powerful tests of H_0 versus H_1.

*Daniel Q. Naiman is Assistant Professor, Department of Mathematical Sciences, Johns Hopkins University, Baltimore, Maryland 21218. This research was supported in part by National Science Foundation Grant DMS-8403646 and by the U.S. Department of the Navy under Office of Naval Research Contract N00014-79-C-0801.

