A Simple Approximation for Calculating Sample Sizes for Comparing Independent Proportions
Author(s): Joseph L. Fleiss, Alex Tytun and Hans K. Ury
Source: *Biometrics*, Vol. 36, No. 2 (Jun., 1980), pp. 343-346
Published by: International Biometric Society
Stable URL: http://www.jstor.org/stable/2529990
Accessed: 10/12/2014 14:38

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
http://www.jstor.org/page/info/about/policies/terms.jsp

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

International Biometric Society is collaborating with JSTOR to digitize, preserve and extend access to *Biometrics.*

http://www.jstor.org

This content downloaded from 128.235.251.160 on Wed, 10 Dec 2014 14:38:51 PM
All use subject to JSTOR Terms and Conditions

BIOMETRICS 36, 343–346
June 1980

A Simple Approximation for Calculating Sample Sizes for Comparing Independent Proportions

Joseph L. Fleiss

Division of Biostatistics, Columbia University School of Public Health, 600 West 168 Street, New York, New York 10032, U.S.A.

Alex Tytun

Office of Biostatistics and Research, New York City Department of Health, 125 Worth Street, New York, New York 10013, U.S.A.

and

Hans K. Ury

Medical Methods Research Department, The Permanente Medical Group, 3700 Broadway, Oakland, California 94611, U.S.A.

SUMMARY

A simple approximation is provided to the formula for the sample sizes needed to detect a difference between two binomial probabilities with specified significance level and power. The formula for equal sample sizes was derived by Casagrande, Pike and Smith (1978, *Biometrics* **34**, 483–486) and can be easily generalized to the case of unequal sample sizes. It is shown that over fairly wide ranges of parameter values and ratios of sample sizes, the percentage error which results from using the approximation is no greater than 1%. The approximation is especially useful for the inverse problem of estimating power when the sample sizes are given.

Suppose a study is being planned to compare two binomial probabilities in independent samples of equal size so that, if α is the significance level for a one-tailed test and if $P_1 < P_2$ are the two underlying probabilities, then 1-β should be the power of the test. Casagrande, Pike and Smith [CPS] (1978) presented an approximate formula for the sample size in each of the two samples required to achieve the desired power.

Define z_p to be the upper $100(1-p)$ percentile of the standard normal distribution, define $\delta = P_2 - P_1$ and define

$$n' = \frac{\{z_\alpha \surd(2\bar{P}\bar{Q}) + z_\beta \surd(P_1Q_1 + P_2Q_2)\}^2}{\delta^2}, \tag{1}$$

where $\bar{P} = \frac{1}{2}(P_1 + P_2)$ and $Q = 1 - P$. Now (1) is the formula for the sample size in each group, that would be derived by analyzing the classic critical ratio test without the continuity correction (see, e.g., Fleiss, 1973). CPS showed that the corrected formula

$$n = \frac{n'}{4}\left\{1 + \sqrt{\left(1 + \frac{4}{n'\delta}\right)}\right\}^2 \tag{2}$$

Key words: Sample sizes; Fourfold tables; Power.

This content downloaded from 128.235.251.160 on Wed, 10 Dec 2014 14:38:51 PM
All use subject to JSTOR Terms and Conditions

for the sample size per group provides an excellent approximation to the sample size obtained by an exact analysis of power (Bennett and Hsu, 1960; Haseman, 1978).

Suppose that considerations of relative cost or other factors make it desirable to select samples of unequal size from the two populations. Let the required sample size from the first population be denoted by m, and that from the second by rm $(0<r<\infty)$, with r specified in advance. The total sample size is, say, $N=(r+1)m$.

As noted by H. K. Ury, in a technical report of the Permanente Medical Group, Oakland, 1978, a simple modification of the CPS development leads to the value

$$m=\frac{m'}{4}\left[1+\sqrt{\left\{1+\frac{2(r+1)}{rm'\delta}\right\}}\right]^2 \tag{3}$$

as the approximate sample size from the first population, and rm as that from the second, which are required to assure a power of $1-\beta$ against the alternative $P_1<P_2$, where

$$m'=\frac{[z_\alpha\sqrt{\{(r+1)\bar{P}\bar{Q}\}}+z_\beta\sqrt{(rP_1Q_1+P_2Q_2)}]^2}{r\delta^2}, \tag{4}$$

$\bar{P}=(P_1+rP_2)/(r+1)$ and $\bar{Q}=1-\bar{P}$. Formula (3) agrees closely with one derived from other principles by Ury. Note that (2) is a special case of (3), and (1) of (4), when $r=1$. The analysis that follows will therefore be for the general case of possibly unequal sample sizes.

To a remarkable degree of accuracy, m is approximately equal to m^*, where

$$m^*=m'+(r+1)/r\delta. \tag{5}$$

Define $x=2(r+1)/rm'\delta$, so that the proportionate difference between m and m^* is, say,

$$R(x)=\frac{m^*-m}{m}=\frac{2+x-2\sqrt{(1+x)}}{2+x+2\sqrt{(1+x)}}.$$

Note that

$$\lim_{x\to 0} R(x)=0, \qquad \lim_{x\to\infty} R(x)=1$$

and $R'(x)>0$ for all $x>0$. Provided that $m'\delta\geqslant 4(r+1)/r$, $x\leqslant 0.50$ and $R(x)\leqslant 0.01$. Thus, for moderately large values of m' (say, $m'\geqslant 120$), moderately large values of δ (say, $\delta\geqslant 0.1$) and sample sizes that are not too disproportionate (say, $0.50\leqslant r\leqslant 2$), the use of the simpler expression in (5) results in a percentage error no greater than 1%.

When one is confronted with the inverse problem of estimating power for prespecified sample sizes, (5) is far simpler to manipulate than (3). Suppose that a one-tailed test with significance level α is to be performed, and suppose also that there is interest in detecting a difference between P_1 and $P_2>P_1$, that N is the available total sample size and that $m^*=N/(r+1)$ is the size of the sample from the first population. Equations (4) and (5) combine to yield

$$z_\beta=\frac{\sqrt{\{r\delta^2m^*-(r+1)\delta\}}-z_\alpha\sqrt{\{(r+1)\bar{P}\bar{Q}\}}}{\sqrt{(rP_1Q_1+P_2Q_2)}} \tag{6}$$

as the approximate percentile corresponding to the actual power. Tables of the normal curve will then provide the power itself. Since only rough estimates of power are usually

This content downloaded from 128.235.251.160 on Wed, 10 Dec 2014 14:38:51 PM
All use subject to JSTOR Terms and Conditions

required, (6) may be used for a wider range of values of r (e.g. $0.33 \leqslant r \leqslant 3$) than that in which (3) and (5) agree well.

Suppose, for example, that $\alpha = 0.05$, that the probabilities $P_1 = 0.15$ and $P_2 = 0.25$ are considered sufficiently different to warrant rejecting the hypothesis of no difference, and that a total sample size of 360 is available, then Table 1 gives the value of z_β from (6) and the corresponding approximate power for several values of r. Note the asymmetry in the table: for example, sample sizes of 270 and 90 from the first and second populations (corresponding to $r = 0.33$) yield an approximate power of 0.63, whereas sample sizes of 90 and 270 (corresponding to $r = 3$) yield an approximate power of 0.58. Other things being equal, power is increased when relatively more observations are taken from the population whose underlying probability is further from 0.50.

Table 1

Approximate powers for detecting a difference between $P_1 = 0.15$ and $P_2 = 0.25$ using a one-sided significance test with a total sample size of 360 and a significance level of 0.05

r†	z_β	Power
0.33	0.32	0.63
0.50	0.49	0.69
1	0.60	0.73
2	0.41	0.66
3	0.19	0.58

† r is the ratio of the sample size from the second population to that from the first.

The power values in Table 1 agree to two decimal places with those obtained by inverting (4) and (3), which must be done by trial and error or iteratively. For more extreme values of r, the discrepancy increases and may be unacceptably large.

If a two-tailed test with significance level α is performed, z_α must be replaced by $z_{\alpha/2}$ and δ must be redefined as $|P_2 - P_1|$.

Acknowledgement

This research was supported in part by a grant from the National Institute of Dental Research.

Résumé

On fournit une approximation simple à la formule nécessaire pour déterminer des tailles d'échantillons destinées à tester la différence entre deux probabilités binomiales avec niveau de signification et puissance fixées. La formule des tailles égales d'échantillon a été obtenue par Casagrande, Pike et Smith (1978, *Biometrics* **34**, 483–486) et elle se généralise facilement au cas de tailles inégales. On montre que pour une large plage des valeurs des paramètres et des rapports de tailles d'échantillon, le pourcentage d'erreur en utilisant l'approximation ne dépasse pas 1%. L'approximation est particulièrement utile pour le problème inverse d'estimation de la puissance quand les d'échantillon sont données.

This content downloaded from 128.235.251.160 on Wed, 10 Dec 2014 14:38:51 PM
All use subject to JSTOR Terms and Conditions

References

Bennett, B. M. and Hsu, P. (1960). On the power function of the exact test for the 2×2 contingency table. *Biometrika* **47**, 393–398.

Casagrande, J. T., Pike, M. C. and Smith, P. G. (1978). An improved approximate formula for calculating sample sizes for comparing two binomial distributions. *Biometrics* **34**, 483–486.

Fleiss, J. L. (1973). *Statistical Methods for Rates and Proportions.* New York: Wiley.

Haseman, J. K. (1978). Exact sample sizes for use with the Fisher–Irwin test for 2×2 tables. *Biometrics* **34**, 106–109.

Received May 1979; *revised July* 1979

This content downloaded from 128.235.251.160 on Wed, 10 Dec 2014 14:38:51 PM
All use subject to JSTOR Terms and Conditions