CORRESPONDENCE

Sample size and power considerations to procure the required number of subjects with a certain attribute

Sir,—Research in epidemiology and clinical medicine is sometimes concerned with deciding on the number of subjects to sample (or the number of subjects without symptoms to screen) to procure the desired number of subjects with attribute X—for example, hypertension. Previously this decision was based solely on the known or postulated prevalence rate of X in the population (π). If π were 0.10 (10 cases of X per 100 persons in the population) and if at least 100 subjects with X were required, the sample size would then be 1000. What this means is that if we took a random sample of 1000 subjects from the population, we would expect 100 such subjects. But this expectation ignores the inevitable sampling error. In fact, there is only a 50% chance that the sample will contain 100 or more subjects with X. A 0.50 power is unacceptably low for estimation of sample size. Clearly we need a statistical procedure to estimate the required sample size (n) to procure at least r subjects with X at a specified power, given that the prevalence rate of X in the population is π, and assuming random sampling. Surprisingly, none of the standard textbooks on biostatistics, epidemiology, or survey methods that I perused consider this important statistical problem in the discussion on sample size and power.

Although the exact solution for this problem (given π, r, and power, find n) is based on the binomial probability distribution, the computations of this exact method are intensive. Therefore I used the normal distribution as an approximation to the binomial distribution.1 Accordingly, we have:

\[ Z_c = \frac{(r - n\pi)}{(n\pi(1-\pi))^{1/2}} \]  

and solve for n. Note that \( Z_c \) is the standard normal deviate corresponding to a specified level of power. Solving for n involves rearranging eq. 1 in the standard form of a quadratic equation (AX^2 + BX + C = 0, where n corresponds to X). After some simple algebra, we obtain:

\[ n = \frac{[Z_c^*\pi(1-\pi)]^{1/2} - (-r)}{\pi\pi(1-\pi)} \]  

By imposing the condition that \( r < n\pi \) or \( Z_c < 0 \), n is obtained by the usual solution of the quadratic equation:

\[ n = \frac{[-Z_c^*\pi(1-\pi)]^{1/2}}{\pi\pi(1-\pi)} \]  

We now consider some examples on the use of equation (3) to find n given π, r, and power: If π = 0.2, r = 40, then n = 200 is required for a power of 0.5, n = 225 for a power of 0.8, n = 252 for a power of 0.95, and n = 264 for a power of 0.975. A computer program coded in BASIC to carry out the computations is listed in the appendix. The program prompts for \( \pi (= \pi) \), r and \( Z_c \) and outputs n. The \( Z_c \) values corresponding to the power levels are given in the one tail standard normal deviate tables found in textbooks on statistics. The \( Z_c \) values for a few selected powers are given below:

<table>
<thead>
<tr>
<th>Power</th>
<th>0.500</th>
<th>0.800</th>
<th>0.900</th>
<th>0.950</th>
<th>0.975</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Z_c )</td>
<td>0.0</td>
<td>0.842</td>
<td>1.282</td>
<td>1.645</td>
<td>1.960</td>
</tr>
</tbody>
</table>

The executable file of the BASIC program is available on request. Please send a diskette for storage.

JAMES LEE
Division of Biostatistics and Health Informatics, Department of Community, Occupational and Family Medicine, National University of Singapore, NUH, Lower Kent Ridge Road, Singapore 0511


Appendix: Listing of the programme coded in BASIC

CLS
COLOR 2, 0
REM Program size.bas [James Lee 1993]
PRINT "Given Pi, r and Zc, find n"
PRINT "--- --- --- --- --- --- ---"
10 PRINT
INPUT "Population prevalence rate of X (Pi)"; Pi
INPUT "Wanted number of subjects in sample with X (r)"; r
INPUT "Z value for desired power (Zc)"; Zc
B = (-Zc) * (Pi * (1 - Pi)) ^ .5
W = (B ^ 2 - (4 * Pi * (-r))) ^ .5
n = (((-B) + W)/(2 * Pi)) ^ 2
PRINT "You need to sample or screen"; n; "(n) subjects"
PRINT "Another calculation (y/n)"
INPUT reply$
IF reply$ = "Y" OR reply$ = "y" THEN GOTO 10
END

Downloaded from http://oem.bmj.com/ on July 7, 2017 - Published by group.bmj.com
Sample size and power considerations to procure the required number of subjects with a certain attribute.

J Lee

Br J Ind Med 1993 50: 765
doi: 10.1136/oem.50.8.765

Updated information and services can be found at:
http://oem.bmj.com/content/50/8/765.citation

These include:

Email alerting service
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/