<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Error in sample size formula</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>Campbell, DD</td>
</tr>
<tr>
<td><strong>Citation</strong></td>
<td>Journal of Human Reproductive Sciences, 2014, v. 7 n. 2, p. 155-156</td>
</tr>
<tr>
<td><strong>Issued Date</strong></td>
<td>2014</td>
</tr>
<tr>
<td><strong>URL</strong></td>
<td><a href="http://hdl.handle.net/10722/203357">http://hdl.handle.net/10722/203357</a></td>
</tr>
<tr>
<td><strong>Rights</strong></td>
<td>This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.</td>
</tr>
</tbody>
</table>


Error in sample size formula

Sir,

Although informative and useful Suresh and Chandrashekara’s article on sample size estimation and power analysis contains a serious error (Suresh and Chandrashekara, 2012). In the section titled “sample size estimation with two means” they state the minimum required sample size for detecting a mean difference between two groups is:

\[ N = \frac{(r+1)(Z_{\alpha/2} + Z_{1-\beta})^2 \sigma^2}{\alpha^2 r d^2} \]

Where
- \( \alpha \) is the false positive rate
- \( \beta \) is the false negative rate
- \( N \) is the sample size required to detect an inter-group mean difference of \( d \) with specified \( \alpha \) and power of 1-\( \beta \)
- \( \sigma^2 \) is the variance in each group (both groups having the same variance)

The formula as stated cannot be correct as relabeling of the two groups results in different values of \( N \).

Example:
If \( n_1 = 100 \) and \( n_2 = 200 \) then \( r = 1/2 \) and \( \frac{(r+1)}{r} = 3 \)
Swapping the two groups around:
Then \( n_1 = 200 \) and \( n_2 = 100 \) then \( r = 2 \) and \( \frac{(r+1)}{r} = 1.5 \)
The \( N \) calculated for the first case is twice that of the second; they should be identical.

In fact, the formula given is the formula for \( n_2 \) which I prove thus.

In the ‘Sample Size estimation with two means’ case, the z-score of the test statistic \( d \) is related to the required false positive rate and power by

\[ \frac{d}{\text{stdErr}(d)} = Z_{\alpha/2} + Z_{1-\beta} \]

Where the standard error of \( d \) is

\[ \text{stdErr}(d) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{\sigma_1^2}{rn_1} + \frac{\sigma_2^2}{rn_2}} = \sqrt{\frac{(1+r)\sigma^2}{rn_2}} \]

Substituting the expression for stdErr \( (d) \) into the first equation and rearranging gives:

\[ n_2 = \frac{(r+1)(Z_{\alpha/2} + Z_{1-\beta})^2 \sigma^2}{\alpha^2 r d^2} \]

The formula for \( N \) is then

\[ N = n_1 + n_2 = rn_1 + n_2 = (r+1)n_2 \]

\[ N = \frac{(r+1)^2 (Z_{\alpha/2} + Z_{1-\beta})^2 \sigma^2}{\alpha^2 r d^2} \]

With the new formula group labels can be swapped without changing the value calculate for \( N \).

The original erroneous formula could result in studies seriously underestimating their required sample size. For instance, the required sample size (as calculated by the current formula) is half that truly required, given equal numbers in the two groups. I therefore draw this error to your attention. The illustrative
examples that follow the formula presentation are also in error.

Desmond Dedalus Campbell\textsuperscript{1,2}
\textsuperscript{1}Department of Psychiatry,\newline\textsuperscript{2}Centre for Genomic Sciences,\newlineUniversity of Hong Kong,\newlineHong Kong, China

Address for correspondence:
Dr. Desmond Dedalus Campbell,\newlineL10-64 Lab Block, Faculty of Medicine,\newline21 Sassoon Road, Central and Western,\newlineHong Kong, China.\newlineE-mail: ddc123@hku.hk

REFERENCE


<table>
<thead>
<tr>
<th>Quick Response Code:</th>
<th>Website:</th>
<th>DOI:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><a href="http://www.jhrsonline.org">www.jhrsonline.org</a></td>
<td>10.4103/0974-1208.138878</td>
</tr>
</tbody>
</table>
Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.