Print ISSN: 2319-2003 | Online ISSN: 2279-0780

IJBCP International Journal of Basic & Clinical Pharmacology

doi: 10.5455/2319-2003.ijbcp20130119

Educational Forum

Sample size: from formulae to concepts - II

Rakesh R. Pathak*

Department of Pharmacology, C.U. Shah Medical College, Surendranagar 363001, India

Received: 10 January 2013
Accepted: 22 January 2013

***Correspondence to:**
Dr. Rakesh R. Pathak,
Email: rr_pathak@yahoo.com

ABSTRACT

Sample size formulae need some input data or to say it otherwise we need some parameters to calculate sample size. This second part on the formula explanation gives ideas of Z, population size, precision of error, standard deviation, contingency etc which influence sample size.

Keywords: Z, Z_β, $Z_{\alpha/2}$, Population size, Contingency

In the last issue[1], we saw how and why we need formulae to calculate the sample size, though there are free online calculators and tables to do this "seeming small" job. In this issue we would see more elaborations on the formulae for sample size calculations. The formula of standard margin of error i.e. Z* $[p*(1-p)/n]^{1/2}$ in the last issue contains standard normal deviate – here indicated by Z which can be obtained from any cumulative normal probability table for different confidence levels. Many books mention "1 – p" as q and sample size 'n' is represented differently as 'e' or 'c'.

In the formula for the standard margin of error, we can calculate the "Z" for 95% confidence limit (1.96 as we got in the last issue) or lower value of 90% (1.645) or higher values like 99% (2.576) or 99.9% (3.29) – though 95% is the most common confidence level. Secondly, we noted that increase in population size beyond 20,000 hardly makes any significant difference in the sample size. Now we would see how this assumption came into being.

Here it must be noted that margin of error is just half of the "confidence interval width which is symmetrically spread across the target ideal value". For example 2.5% margin of error in a 100 gm ideal weight means finally values should be between 97.5 gm and 102.5 gm, thus (102.5 – 97.5) = 5% would be the total confidence interval.[2]

For 1% margin of error and 95% confidence level the maximal sample size (at p = 50%, thus any sampled item having equal chances of giving positive or negative result) the calculated sample size was 9600 (9604 to be more precise). For an allowed percentage error of 5%, the sample size is reduced to 385 and this sample size is assumed for an infinite population.

In case the population size (from which sample is taken) is known, we can calculate its effect on the sample size. The formula is SS/ [1 + {(SS – 1)/ PS}] where SS = sample size and PS = population size. Thus for a population of 20,000 and 5% permitted error, the SS = 377 instead of 385 i.e. a difference of < (–2)%.

But this reduction in sample size is more effective when we calculate for more precision i.e. 1% – then the sample size for infinite population is large i.e. 9604 and for a population of 20,000, it would come down to 6489 – nearly one third reduction in the sample size.

It means that we need not increase sample size as rapidly as the size of parent population increases and this fact is more true for higher precision levels.[3] A simpler formula for calculating the sample size is also available – SS = PS / [1 + (PS *PE^2)] where SS = sample size, PS = population size and PE = precision of error (if 5% error permitted, PE written as 0.05).[4]

For a population of 20,000 and permitted error of 0.05, the required size = 392. If the precision is 0.01, the sample size from this formula would be 10,000 and both these values are slightly more than our previously calculated values of 385 & 9604 (to adjust for contingency that we would review later in this article).

But due to other reasons, sample size might need an increase over a calculated value. For example, if we are using cluster sample (like anthropometric values in a village survey) instead of simple random sampling, we usually multiply the sample size by 2 and add extra 5% as contingency (non-response or recording error). 30 is the standard number of clusters established by the WHO Expanded Program of Immunization (EPI Cluster Surveys).[5]

For just estimating mean value, the sample size can also be calculated from a previously known standard deviation and pre-determined margin of error. For example, if 5% is the acceptable margin of error and 15 is the known standard deviation, $4SD^2/ME^2 = 4*15*15/(5*5) = 36$ would be the required sample size.[2]

For a pre-set mean difference = MD that must be detectable (suppose 1 unit) while comparing two samples, we can find sample size with a given standard deviation. For example, if the standard deviation = SD = 15, $[16 * SD^2/ (MD)^2] + 1= [16* 900/ 1] + 1 = 3600$ would be the required sample size at 80% power.[2]

Here the constant value of 16 in the formula is a round off value. For 80% power (which is most commonly used), $Z_\beta = 0.84$ and for 95% significance level $Z_{\alpha/2} = 1.96$. Putting both the values in the formula $SS = [2\sigma^2 *(Z_\beta + Z_{\alpha/2})^2] / MD^2$ where MD = mean difference, {in many books, MD is also called effect size = ES} we get a computed multiplier of 15.7 which is rounded off to 16 as said earlier.[6]

For more advanced knowledge in sample size determination according to study design (e.g. superiority trial, non-inferiority trial etc) one can refer an excellent article by Zhong B.[7]

REFERENCES

1. Pathak RR. Sample size: from formulae to concepts. Int J Basic Clin Pharmacol 2012;1:223-4.
2. Sample Size, Precision, and Power. Available at http://www.sjsu.edu/faculty/gerstman/StatPrimer/sampsize.PDF. Accessed 9 December 2012.
3. Israel GD. Determining Sample Size. Available at http://edis.ifas.ufl.edu/pd006. Accessed 7 December 2012.
4. Yamane T. Statistics: An Introductory Analysis, 2nd Ed. New York: Harper and Row; 1967.
5. Calculating the Sample Size. Available at http://www.ifad.org/gender/tools/hfs/anthropometry/ant_3.htm. Accessed 15 December 2012.
6. Introduction to sample size and power calculations. Available at http://www.stanford.edu/~kcobb/hrp259/lecture11.ppt. Accessed 15 December 2012.
7. Zhong B. How to Calculate Sample Size in Randomized Controlled Trial? J Thorac Dis 2009;1:51-4.

doi: 10.5455/2319-2003.ijbcp20130119
Cite this article as: Pathak RR. Sample size: from formulae to concepts - II. Int J Basic Clin Pharmacol 2013;2:94-5.