The SAMPLE SIZE

Kumar A[1]

ABSTRACT

Finding an "appropriate sample size" has been the most basic and foremost problem; a research worker is always faced with, in all sampling based analytical researches. This is so, since a very large sized sample results to unnecessary wastage of resources, while a very small sized sample may affect adversely the accuracy of sample estimates and thus in turn losing the very efficacy of selected sampling plan. The present paper attempts to highlight the main determinant factors and the analytical approach towards estimation of required sample size, along with a few illustrations.

KEY WORDS: Sample size, Probability, Margin of error, Probability level, Confidence coefficient, Level of significance, Data variability, Normal distribution

1. Associate Professor of Biostatistics, Department of Community Medicine, Universal College of Medical Sciences & Teaching Hospital, Bhairahawa, Nepal

For Correspondence
Dr.Ashok Kumar
Associate Professor of Biostatistics,
Department of Community Medicine
Universal College of Medical Sciences & Teaching Hospital,
Bhairahawa, Nepal
Email: aku001@rediffmail.com

INTRODUCTION

What should be the size of a sample is a very pertinent question every research worker is coming across in all types of researches. Determining appropriate sample size i.e. the number of units to be selected in the sample, has always been a key problem in sampling theory. The simple reason is that (1) a very large sized sample results to unnecessarily wastage of resources, while (2) a sample of very small size may affect adversely the accuracy of sample estimates and thus in turn the very utility of sampling plan.

DETERMINANT FACTORS

Apart from various constraints in terms of available resources, like those of cost, time, man power, and material requirements; the decision making in respect of "size of sample" is mainly governed by following three factors:

(1) Tolerable margin of error.
(2) Specified probability level.
(3) The data variability.

It may be mentioned at this point, that subject to acceptable margin of error in terms of accuracy of sample values; chosen probability level; another important factor playing an important role in determining sample size is the data variability. If, the data variability is high the sample has to be of large size; while in case of low data variability a sample can be very well of small size.

ANALYTICALAPPROACH

The calculation of sample size, corresponding to specified margin of error and the probability level; is based on assumption of normal distribution using tables of normal deviates at specified probability levels. If, mean and Standard Deviation (SD) of a population are respectively μ and σ and the margin of error (deviation of sample mean from actual population mean) is E, then:

(1) At 0.05 probability level (confidence coefficient of 0.95), the sample size(n) is given by:
$\mathbf{n = (1.96\sigma/E)^2}$

(2) At 0.01 probability level (confidence coefficient of 0.99), the sample size (n) is given by:
$\mathbf{n = (2.58\sigma/E)^2}$

It is also to be stated that the relationships (1) and (2) hold good under the assumption that the Population Size (N) is very large and that fpc(finite population correction) can be ignored i.e.(N-n)/N can be taken as 1.

Further, in estimation of Sample Size (n), using above relationship, Population Data Variability i.e. SD (σ) is in general not known; so in such cases first we estimate it on the basis of a sample of arbitrary size through pilot sample survey and then proceed to determine the actual sample size.

ILLUSTRATIONS

A few illustrations; in the form of PROBLEM-SOLUTION in calculation of REQUIRED SAMPLE SIZE, are presented as under:

PROBLEM 1: In a population, mean diastolic blood pressure is stated to be 85 mmHg with SD of 8 mmHg. Calculate the minimum size of sample required verifying the above: subject to margin of error of 2 mmHg and a risk percentage of 5%.

Solution: Given that,
Population Standard Deviation (σ)= 8 mmHg
Margin of Error (E) = $\pm$ 2 mmHg
Risk Percentage = 5%

Therefore, minimum size of sample (n) at 5% Risk Percentage (0.05 Probability Level) will be given by:
$n = (1.96\sigma/E)^2 = (1.96x8/2)^2 = (7.84)^2 = 61.46$ i.e 61

PROBLEM 2: The Mean Pulse Rate in a Population is estimated as 75 beats per minute with SD of 10 beats. What should be the minimum size of sample to test the validity of above, at risk percentage of 5% and a margin of Error of 2 beats?

Solution: Given that,
Population SD (σ) = 10 Beats per Minute
Margin of Error (E) = $\pm$ 2 Beats
Risk Percentage = 5%

Therefore, minimum size of sample (n) at 5% risk percentage (0.05 Probability Level) will be given by:
$n = (1.96\sigma/E)^2 = (1.96x10/2)^2 = (9.80)^2 = 96.04$ i.e 96

SAMPLE SIZE IN QUALITATIVE CHARACTERS

The Sample Size in case of Qualitative Characters (attributes), like Incidence Rate, Prevalence Rate, Morbidity Rate, and Cure Rate etc. is worked out on the basis of proportion of the units possessing the specified attribute, to the total number of units in the population; in the usual manner as already outlined.

In such cases, if
p = a positive integer = proportion of units possessing the attribute under Study to the total number of units in the population.

E = Tolerable margin of error (Under the general assumption that It is not more than 10% to 20% of p).

Then, the minimum size of sample (n) at 5% riskpercentage (0.05 Probability Level) is given by:

$n = (1.96\sqrt{pq}/E)^2 = (3.84)pq/E^2$
Where, $q = 1 - p$.

PROBLEM: In a population of a specific area the incidence rate of a disease in an epidemic was recorded as 15 per 100. Find the minimum size of the sample required to workout incidence rate of the same disease in that area in the current epidemic; subject to permissible margin of error of 10%.

Solution: Given that,
p = incidence rate of the disease in the last epidemic.
= 15 per 100 = 15/100 = 0.15
Therefore, q = (1-p) = 1 0.15 = 0.85.

Permissible Margin of Error (E) = 10% of p =10% of 0.15 = 0.015.

Therefore, minimum size of sample at 5% Risk Percentage (0.05 Probability Level) will be given by:
$n = (3.84)/PQ/E^2 = (3.84)(0.15)(0.85)/(0.015)^2 = 2176$.

CONCLUSION

It is to be pointed out here, that no doubt, sample size is determined, mainly on the basis of specified probability level and tolerable margin of error. But, never the less the same can also be decided as per "feasibility"and in such cases we can reduce sample size accordingly by increasing permissible margin of error, and as such ALTERNATIVE SAMPLE SIZES can be worked out by varying, probability level and the margin of error.

REFERENCES

1. *Cochran WG. Sampling techniques 3rd ed. New York: John Wiley & Sons; 2007.*
2. *Hansen MH, Hurwitz WN, Madow WG. Sampling survey methods and theory vol. 1. New York: John Wiley & Sons; 1953. PMid:13087864*
3. *Kumar A. An introduction to sampling. Allahabad: Students Friends Publications; 2004.*
4. *Mahajan BK. Methods in biostatistics. 7th ed. New Delhi: JP; 2010.*
5. *Pant PR. Social science research and thesis writing. 6th ed. kathmandu: Buddha Academic Publishers; 2012*
6. *Sukhatme PV, Sukhatme BV, Sukhatme S, Ashok C. Sampling theory of surveys with applications: Iowa State University Press and Indian Society of Agricultural Statistics, New Delhi; 1984. PMid:6440142 PMCid:PMC392217*