Sample Size

The size of the sample is an important factor. It has direct bearing on the accuracy, estimation, cost and administration of the survey. Large sample has low sampling error where as small sample have higher sampling error. To avoid unnecessary cost small sample should be selected. Hence optimum sample size should be selected to fulfill the requirement of efficiency, representativeness, reliability and flexibility. Some of the factors affecting the sample size are nature of study, nature of reaction of respondent towards the subject under study, nature of population i.e. composition of population under study, number of classes in the population, types of sampling used during study etc.

Factors affecting Sample Size

Size of sample depends upon different factors. These are

i. Nature of population
ii. Number of classes
iii. Nature of the study
iv. Types of sampling used
v. Degree of accuracy

Nature of population

If the population under study is homogeneous then small sample size is sufficient, but in case of heterogeneous population large sample size is required to make sample size representative of the population.

Number of classes

For the classification with large number of classes, large sample size is required.

Nature of study

If the study takes long time then small sample size is better from the financial and analysis point of view.

Types of sampling used

The sample size depends upon the type of sampling used. For simple random sampling large sample size is required but for the case of stratified sampling small sample size is sufficient.

Degree of accuracy

If the greater degree of accuracy is required then large sample should be selected.

Testing Reliability of the Sample

If the selected sample is representative of the population then sample is called reliable. The selected sample is reliable or not can be tested using following methods;

i. Drawing parallel sample.
ii. Comparing sample with population.
iii. Drawing sub sample from main sample.

Drawing parallel sample

Draw sample parallel to the drawn sample from the population and compare various measures such as average, dispersion, skewness, kurtosis etc. between the samples. If the comparison measures are alike then the sample is reliable otherwise unreliable.

Comparing sample with population

Different measures computed from samples are compared with that of population. If the measures are identical then the selected sample is reliable.

Drawing sub sample from main sample

The different measures computed from sub sample arc compared with main sample. It cannot be used to find the sample is representative of population or nut but can be used to find if any error occurred due to faulty selection of sample.

Method of Estimating Sample Size

Estimation of sample size by using mean

Let $\bar{x}$ be the sample mean from a random sample of size n drawn from population with mean E($\bar{x}$) and standard deviation σ.

Now,

$$Z = \frac{\bar{x} - E(\bar{x})}{SE(\bar{x})} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

at α level of significance and $(1-\alpha)$ confidence limit is

$$P\left(\left|\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}\right| \leq Z\alpha/_2\right) = 1 - \alpha$$

$P(|\bar{x} - \mu| \le \sigma/\sqrt{n})\ Z_{(}\alpha/2) = 1 - \alpha$

Now, x - µ = d (margin of error) then

d = $\frac{\sigma}{\sqrt{n}} Z\, \alpha/2$

$\sqrt{n} = \frac{\sigma}{d} Z\alpha/2$

n = $\frac{\sigma^2\ Z_{\alpha/2}{}^2}{d^2}$

In case of σ is not known take $\sigma = s$

For the finite population of Size N, sample size = $\frac{\sigma^2\ Z_{\alpha_2}{}^2}{d^2 + \frac{\sigma^2\ Z_{\alpha_2}{}^2}{N}} = \frac{n}{1+\frac{n}{N}}$

Estimation of sample size by using proportion

Let p be sample proportion from random sample of size n drawn from population with proportion P

Now,

$Z = \frac{p-E(p)}{SE(p)} = \frac{p-P}{\sqrt{\frac{PQ}{n}}}$

At α level of significance (1 - α) confidence limit is

$$P\left(\left|\frac{p-P}{\sqrt{\frac{PQ}{n}}}\right| \le Z\alpha/_2\right) = 1 - \alpha$$

or P $(|p - P| \le \sqrt{\frac{PQ}{n}}\ Z\alpha/_2$ = 1 - α

Now, p - P = d (margin of error) then

d= $\sqrt{\frac{PQ}{n}}\ Z\alpha/_2$

$\sqrt{n} = \frac{Z_{\alpha/2} \sqrt{PQ}}{d}$

n= $\frac{PQ\ Z_{\alpha/2}{}^{2}}{d^2}$

In case of P is not known take P = p.

For the finite population of size N, sample size $= \frac{PQ\ Z_{\alpha/2}{}^{2}}{d^2 + \frac{PQ\ Z_{\alpha/2}{}^{2}}{N}} = \frac{n}{1+\frac{n}{N}}$

Example 1

Determine the minimum sample size required so that the sample estimate lies within 10% of the true value with 95% level of confidence when coefficient of variation is 60%.

Solution

Here,

C.V. = 60% = 0.6

P $(|\bar{x} - \mu| \leq 0.1\ \mu\ )$ = 0.95

(i) Confidence level (1 - α) = 95% = 0.95 then α = 0.05

Now,

$P\left(\left|\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}\right| \leq Z_{\alpha/2}\right) = 1 - \alpha$

$P(|\bar{x} - \mu| \leq \sigma/\sqrt{n})\ Z_{(}\alpha/2) = 0.95$

$P(|\bar{x} - \mu| \leq 1.96 \times \frac{\sigma}{\sqrt{n}}) = 0.95$

From equation (i) and (ii)

$0.1\mu = 1.96 \times \frac{\sigma}{\sqrt{n}}$

$\sqrt{n}= \frac{1.96}{0.1} \times \frac{\sigma}{\sqrt{n}}$

n= $\left(\frac{1.96}{0.1} \times \frac{\sigma}{\sqrt{n}}\right)^2$

n = 384.16 × CV^2

n = 384.16 × $(0.6)^2$

or n = 138.29 = 138

Hence required sample size is 138.

Example 2 In measuring reactions time, a psychologist estimates that the standard deviation is 0.05 seconds. How large a sample of measurement must be taken in order to be 99% confident that the error of his estimate will not exceed 0.01 seconds?

Solution

Here,

Sample size (n) = ?

Standard deviation (s) = 0.05

Confidence interval (1 - α) = 99% = 0.99

or α = 0.01 $Z_{\alpha/2}$= 2.58

Error (d) = 0.01

Here σ = s

n = $\frac{\sigma^2 \ {Z_{\alpha_2}}^2}{d^2} = \frac{(0.05)^2(1.96)^2}{(0.01)^2}$ = 166.4 ≈ 167

Hence required sample size is 167.

Example 3

A researcher wants to conduct a survey of disabled at Kathmandu valley. What should be the sample size of the prior estimate of population of disables if the population is 10% and the desired error of estimation is 2% and level of significance is 5%.

Solution

Sample size (n) = ?

Population proportion (p) = 10% = 0.1

q = 1 - p = 0.9

Error (d) = 2% = 0.02

Level of significance (α) = 5%

Here P = p

$$n= \frac{PQ\, {Z\alpha/_2}^2}{d^2} = \frac{(1.96)^2 \times 0.1 \times 0.9}{(0.02)^2} = 864.36 \approx 865$$

Hence required sample size is 865**.**

Example 4

For p 0.2, d = 0.05 and z = 2 find n. Also find n if N = 1000.

Solution

Here P= p

Now

$$n= \frac{PQ\, {Z\alpha/_2}^2}{d^2} = \frac{4 \times 0.2 \times (1-0.2)}{0.05^2} =256$$

When N = 1000

$$\text{Sample size} = \frac{n}{1+\frac{n}{N}} = \frac{256}{1+\frac{256}{1000}} = 203.82 \approx 204$$

Example 5

The mean systolic blood pressure of a certain group of people was found to be 125 mm of Hg with standard deviation of 15 mm of Hg. Calculate sample size to verify the result at 5'y level of significance if error do not exceed 2. Also find sample size if sample is selected from population of size 500.

Solution

Standard deviation (s) = 15

Level of significance (α) = 5%

Sample size (n) = ?

Error (d) = 2 Here σ = s

Now,

$$n = \frac{\sigma^2 \ {Z_{\alpha_2}}^2}{d^2} = \frac{(15)^2(1.96)^2}{(2)^2} = 216.09 \approx 216$$

When N = 500

$$\text{Sample size} = \frac{n}{1+\frac{n}{N}} = \frac{216}{1+\frac{216}{500}} = 150.83 \approx 151$$

YAMANE FORMULA

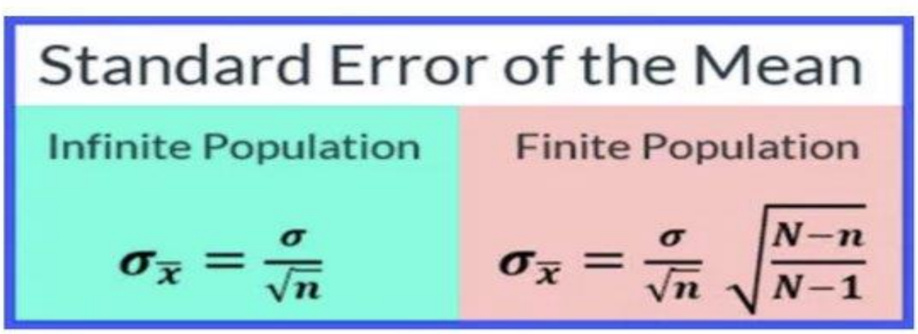
$$n = \frac{N}{1+N \times e^2}$$

Where n= the sample size

N= the population size

e= the acceptable sampling error

95% confidence level and p= 0.5 are assumed



Working with a finite population and if the population size is known, the **Yamane formula for determining the sample size is given by:**

$n=N/(1+N e^2)$

Where

n= corrected sample size, N = population size, and e = Margin of error (MoE), e = 0.05 based on the research condition.

Let's assume that the population is 10,000. At 5% MoE., the sample size would be: $10000/(1+10000(.05^2)$

$=10000/26$

$= 384.61 \sim 385$

In a finite population, when the original sample collected is more than 5% of the population size, the corrected sample size is determined by using the Yamane's formula.

In the example above, 5% of 10,000 is 500 and hence the corrected size is 385 although for research purposes, even 385 is a big number (for handling and collection point of view) and the researcher has to make a decision to collect even smaller number in order of ease of handling, costing but he has to ensure that the sample is representative.