Data Science 1
Statistical Inference
Estimation for One Population
Sample Size Determination

Ann Maharaj

1. Introduction
2. Sample Size Determination - Population Mean
3. Sample Size Determination - Population Proportion
4. Sample Size Determination - Population Variance
5. Factors Affecting Sample Size
6. Summary

Introduction

- We will determine the appropriate sample size to estimate the population mean, proportion and variance.
- In doing so, the following factors which affect the sample size need to be addressed:
  1. How precise do the estimates obtained from the sample need to be?
     - Required Level of Precision
  2. How confident do you wish to be that your results represent what is going on in the population?
     - Required Level of Confidence
  3. How much variability exists in the population?
     - Preliminary Estimate of the Population Standard Deviation or Population Proportion
  4. What cost is involved?
     - How large or small a budget is available?

Introduction

- We will address the first three factors after we work through examples to determine the samples sizes for estimating the population parameters.
- We will address the factor of how cost affect the sample size through an example at the end.

Sample size determination for estimating the population mean

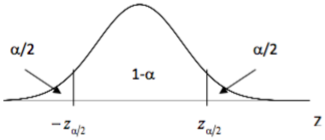
$$P\left(-z_{\alpha/2} < Z < z_{\alpha/2}\right) = 1-\alpha$$

$$\Rightarrow P\left(-z_{\alpha/2} < \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < z_{\alpha/2}\right) = 1-\alpha$$

$$\Rightarrow P\left(-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \overline{X}-\mu < z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

$$\Rightarrow P\left(|\overline{X}-\mu| < z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

$$Z = \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$



So that if the sampling error is to be less than E, then

$$E = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

E is the **maximum acceptable sampling error** for the level of confidence $1-\alpha$

Sample size determination for estimating the population mean

- For an *infinite* population

$$E = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Solving for n, we get

$$n = \left(\frac{z_{\alpha/2} \times \sigma}{E}\right)^2$$

- If the population is finite

$$E = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

Solving for n, we get

$$n = \frac{N(z_{\alpha/2} \times \sigma)^2}{(N-1)E^2 + (z_{\alpha/2} \times \sigma)^2}$$

Sample size determination for estimating the population mean

- If σ is unknown, then a preliminary estimate σ_{est} is used.

- For an *infinite* population

$$n = \left(\frac{z_{\alpha/2} \times \sigma_{est}}{E} \right)^2$$

- For a finite population

$$n = \frac{N(z_{\alpha/2} \times \sigma_{est})^2}{(N-1)E^2 + (z_{\alpha/2} \times \sigma_{est})^2}$$

Preliminary estimate of σ

- More often than not, when determining the sample size to estimate the population mean, the population standard deviation is unknown and has to be estimated.
- But since we do not have a sample yet, how do we estimate it?
- We obtain a preliminary estimate from historical data or from a small preliminary sample or from the use of the empirical rule.
- The empirical rule which is an *important rule of thumb* is used to state the approximate percentage of values that lie within a given number of standard deviations from the mean for a set of data if the data are values of a variable that is normally distributed.
- The requirement that the data come from a normal distribution contains some tolerance, and the empirical rule generally applies if the distribution of the data is approximately bell-shaped.

Empirical Rule

- If a set of data has an approximate bell-shaped distribution, then approximately
  - 68% of the values are within one standard deviation of the mean.
  - 95% of the values are within two standard deviations of the mean.
  - 100% of the values are within three standard deviations of the mean.

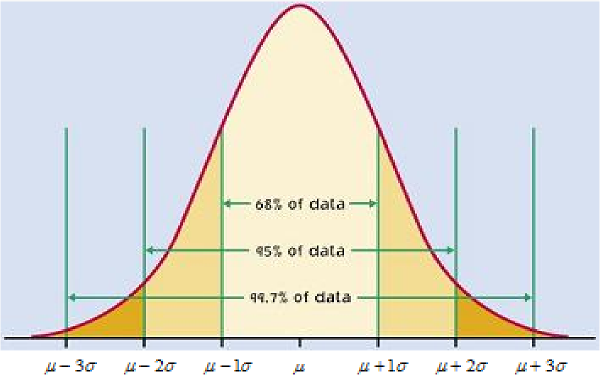


Image from Blogspot

Preliminary estimate of σ

So from the empirical rule, approximately, 100% of observations associated a random variable X lie between -3 and 3 standard deviations from the mean.

$$P(\mu - 3\sigma < X < \mu + 3\sigma) \approx 1$$

$$\mu + 3\sigma - (\mu - 3\sigma) = 6\sigma \approx Range \Rightarrow \sigma \approx \frac{1}{6} Range$$

Therefore, we can use $\frac{1}{6} Range$ as a preliminary estimate of σ.

Example 1

Auditors for the Paxton Plumbing Company want to develop a 95% confidence interval of mean sales invoices with a sampling error of no more than \$5. Past data indicates that the standard deviation of the sales amount has been approximately \$25 for a substantial period of time.

1. What sample size should be taken?
2. Suppose that the company has 500 sales invoices this month. What sample size should be taken?

Example 1: Part 1

1)
$E = 5, \sigma_{est} = 25,$
$(1 - \alpha) = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow z_{\alpha/2} = z_{0.025}$

$$n = \left(\frac{z_{\alpha/2} \times \sigma_{est}}{E} \right)^2$$

We will now compute the sample size in R.

Example 1: Part 1 - R code and solution

```
> #sampling error
> E <- 5
> #prelimary estimate of sigma
> sigma_est <- 25
> #97.5th percentile of the standard normal distribution
> z025 <- qnorm(.975)
> z025
[1] 1.959964
> #sample size
> n <- ((z025*sigma_est)/E)^2
> n
[1] 96.03647
```

Hence, the sample size that should be taken is at least 97.
Note: We always round the sample size UP to the next integer.

Example 1: Part 2

2)
$E = 5, \sigma_{est} = 25, 1 - \alpha = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow z_{\alpha/2} = z_{0.025}, N = 500$

$$n = \frac{N(z_{\alpha/2} \times \sigma_{est})^2}{(N-1)E^2 + (z_{\alpha/2} \times \sigma_{est})^2}$$

We will now compute the sample size in R.

Example 1: Part 2 - R code and solution

```
> #sampling error
> E <- 5
> #prelimary estimate of sigma
> sigma_est <- 25
> #population size
> N <- 500
> #97.5th percentile of the standard normal distribution
> z025 <- qnorm(.975)
> z025
[1] 1.959964
> #sample size
> num <- N*(z025*sigma_est)^2
> dem <- ((N-1)*(E^2))+(z025*sigma_est)^2
> n <- num/dem
> n
[1] 80.69797
```

Hence, the sample size that should be taken is at least 81. Notice, that the required sample size is less when the population size is known, since more information is available.

The impact on sample size by varying components

Consider the sample size

$$n = \left(\frac{z_{\alpha/2} \times \sigma}{E} \right)^2$$

- If σ and E are fixed, as the level of confidence increases
  - $z_{\alpha/2}$ increases; hence the sample size n increases.
- If σ and the level of confidence is fixed, as the precision increases
  - E decreases; hence the sample size increases.
- If the population standard deviation σ is unknown, a preliminary estimate σ_{est} is taken and the levels of confidence and precision are fixed:
  - A larger σ_{est} will result in the sample size n increasing

Varying the level of confidence

$$n = \left(\frac{z_{\alpha/2} \times \sigma}{E} \right)^2$$

σ and E are fixed, e.g., $E = 5, \sigma = 25$.

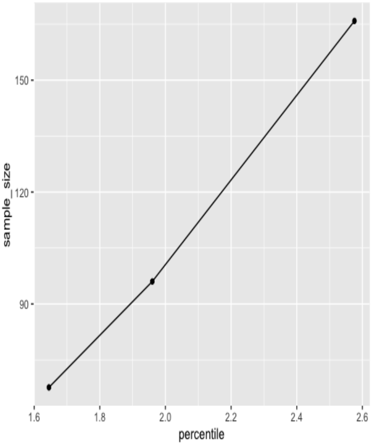
Level of confidence $(1 - \alpha)$ increases

$\Rightarrow$ Percentile, $z_{\alpha/2}$ increases

$\Rightarrow$ Sample size n increases

Table: Varying the level of confidence

$1-\alpha$	α	$z_{\alpha/2}$	n
90%	10%	1.644854	68
95%	5%	1.959964	97
99%	1%	2.575829	166



Varying the level of precision

σ and $(1-\alpha)$ are fixed, e.g., $\sigma = 25$, $(1-\alpha) = 0.95$.

$$n = \left(\frac{z_{\alpha/2} \times \sigma}{E}\right)^2$$

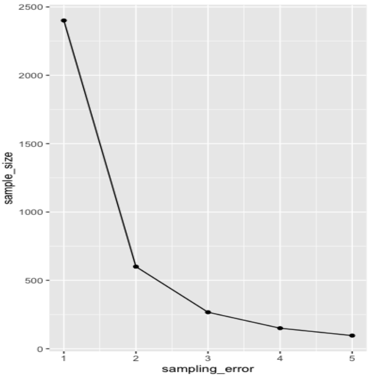
Level of precision increases

$\Rightarrow$ Sampling error, E decreases

$\Rightarrow$ Sample size n increases

Table: Varying the level of precision

Sampling Error (E)	Sample Size (n)
5	97
4	151
3	267
2	601
1	2401



Varying the preliminary estimate of the standard deviation

E and $(1-\alpha)$ are fixed, e.g., $E = 5$, $(1-\alpha) = 0.95$.

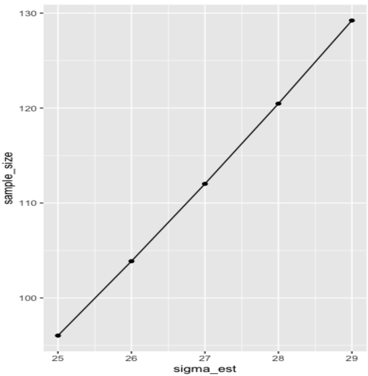
$$n = \left(\frac{z_{\alpha/2} \times \sigma}{E}\right)^2$$

Preliminary estimate of the standard deviation σ_{est} increases

$\Rightarrow$ Sample size n increases

Table: Varying the preliminary estimate of the standard deviation

σ_{est}	Sample Size (n)
25	97
26	104
27	113
28	121
29	130



Sample size determination for estimating the population proportion

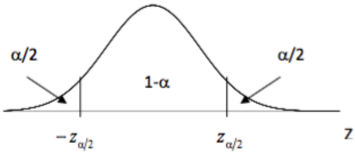
$$P\left(-z_{\alpha/2} < Z < z_{\alpha/2}\right) = 1 - \alpha$$

$$Z = \frac{p - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} \sim N(0,1)$$

$$\Rightarrow P\left(-z_{\alpha/2} < \frac{p - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} < z_{\alpha/2}\right) = 1 - \alpha$$

$$\Rightarrow P\left(-z_{\alpha/2}\sqrt{\frac{\pi(1-\pi)}{n}} < p - \pi < z_{\alpha/2}\sqrt{\frac{\pi(1-\pi)}{n}}\right) = 1-\alpha$$

$$\Rightarrow P\left(|p - \pi| < z_{\alpha/2}\sqrt{\frac{\pi(1-\pi)}{n}}\right) = 1 - \alpha$$



So that if the sampling error is to be less than E,

then

$$E = z_{\alpha/2}\sqrt{\frac{\pi(1-\pi)}{n}}$$

E is the **maximum acceptable sampling error** for the level of confidence $1 - \alpha$.

Sample size determination for estimating the population proportion

- For an *infinite* population

$$E = z_{\alpha/2}\sqrt{\frac{\pi(1-\pi)}{n}}$$

Solving for n, we get

$$n = \left(\frac{z_{\alpha/2} \times \sqrt{\pi(1-\pi)}}{E}\right)^2$$

- If the population is finite

$$E = z_{\alpha/2}\sqrt{\frac{\pi(1-\pi)}{n}}\sqrt{\frac{N-n}{N-1}}$$

Solving for n, we get

$$n = \frac{n_0}{1 + \frac{n_0 - 1}{N}}$$

$$n_0 = \left(\frac{z_{\alpha/2} \times \sqrt{\pi(1-\pi)}}{E}\right)^2$$

Sample size determination for estimating the population proportion

- If the population proportion π is unknown, a preliminary estimate π_{est} is used
  - For an *infinite* population

$$n = \left(\frac{z_{\alpha/2} \times \sqrt{\pi_{est}(1 - \pi_{est})}}{E} \right)^2$$

  - For a infinite population

$$n_0 = \left(\frac{z_{\alpha/2} \times \sqrt{\pi_{est}(1 - \pi_{est})}}{E} \right)^2$$

- If there is no preliminary estimate of proportion, then a value of n can be obtained by using $\pi_{est} = 0.5$.

Example 2

Suppose that we wish to obtain a 95% confidence interval estimate of the proportion of voters who choose a particular election candidate.

1. How many voters should be polled if we want the precision to be 0.01 or less.

2. Assume now that the population of voters in this seat is known to be 12,350. Given the same conditions described in Part (1), calculate the sample size.

Example 2: Part 1

1)
$E = 0.01, \pi_{est} = 0.5,$
$1 - \alpha = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow z_{\alpha/2} = z_{0.025}$

$$n = \left(\frac{z_{\alpha/2} \times \sqrt{\pi_{est}(1 - \pi_{est})}}{E} \right)^2$$

We will now compute the sample size in R.

Example 2: Part 1 - R code and solution

```
> #sampling error
> E <- 0.01
> #prelimary estimate of pi
> pi_est <- 0.5
> #97.5th percentile of the standard normal distribution
> z025 <- qnorm(.975)
> z025
[1] 1.959964
> #sample size
> n <- (z025*(sqrt(pi_est*(1-pi_est)))/E)^2
> n
[1] 9603.647
```

Hence, the sample size that should be taken is at least 9604.
Note: We always round the sample size UP to the next integer.

Example 2: Part 2

2)
$E = 0.01, \pi_{est} = 0.5, N = 12,350,$
$1 - \alpha = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow z_{\alpha/2} = z_{0.025}$

$$n = \frac{n_0}{1 + \frac{n_0 - 1}{N}}$$

$$n_0 = \left(\frac{z_{\alpha/2} \times \sqrt{\pi(1-\pi)}}{E} \right)^2$$

We will now compute the sample size in R.

Example 2: Part 3 - R code and solution

```
> #sampling error
> E <- 0.01
> #prelimary estimate of pi
> pi_est <- 0.5
> #population size
> N <- 12350
> #97.5th percentile of the standard normal distribution
> z025 <- qnorm(.975)
> z025
[1] 1.959964
> #sample size
> n0 <- (z025*(sqrt(pi_est*(1-pi_est)))/E)^2
> n0
[1] 9603.647
> n <- n0/(1+(n0-1)/N)
> n
[1] 5402.767
```

Hence, the sample size that should be taken is at least 5403.
The required sample size is less when the population size is known, since more information is available.

Impact on sample size by varying components

Consider the sample size

$$n = \left(\frac{z_{\alpha/2} \times \sqrt{\pi(1-\pi)}}{E} \right)^2$$

- If π and E are fixed, as the level of confidence increases
  - $z_{\alpha/2}$ increases; hence the sample size n increases.
- If π and the level of confidence is fixed, as the precision increases
  - E decreases; hence the sample size increases.
- If the population proportion π is unknown, a preliminary estimate π_{est} is taken and the levels of confidence and precision are fixed:
  - A smaller π_{est} will result in the sample size n increasing.

Varying the level of confidence

$$n = \left(\frac{z_{\alpha/2} \times \sqrt{\pi(1-\pi)}}{E} \right)^2$$

π and E are fixed, e.g., $E = 0.01, \pi = 0.5$.

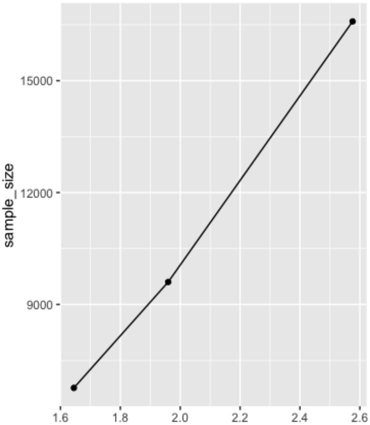
Level of confidence $(1-\alpha)$ increases

$\Rightarrow$ Percentile, $z_{\alpha/2}$ increases

$\Rightarrow$ Sample size n increases

Table: Varying the level of confidence

$1-\alpha$	α	$z_{\alpha/2}$	n
90%	10%	1.644854	6764
95%	5%	1.959964	9604
99%	1%	2.575829	16588



Varying the level of precision

π and $(1-\alpha)$ are fixed, e.g., $\pi = 0.5$, $(1-\alpha) = 0.95$

$$n = \left(\frac{z_{\alpha/2} \times \sqrt{\pi(1-\pi)}}{E} \right)^2$$

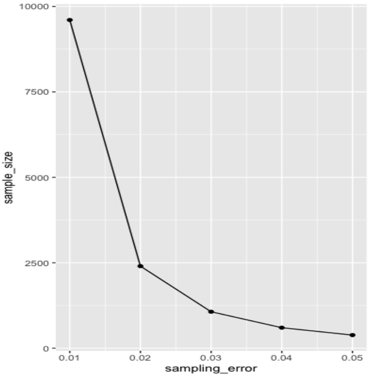
Level of precision increases

$\Rightarrow$ Sampling error, E decreases

$\Rightarrow$ Sample size n increases

Table: Varying the level of precision

Sampling Error (E)	Sample Size (n)
0.05	385
0.04	601
0.03	1068
0.02	2401
0.01	9604



Varying the preliminary estimate of the proportion

E and $(1-\alpha)$ are fixed, e.g., $E = 0.01$, $(1-\alpha) = 0.95$

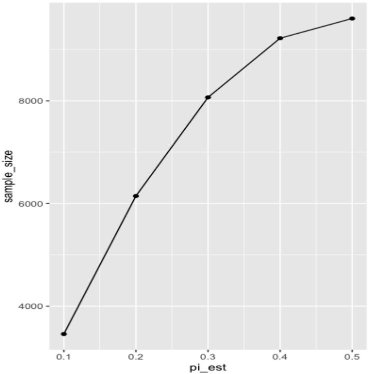
$$n = \left(\frac{z_{\alpha/2} \times \sqrt{\pi_{est}(1-\pi_{est})}}{E} \right)^2$$

Preliminary estimate of the proportion π_{est} increases

$\Rightarrow$ Sample size n increases

Table: Varying the preliminary estimate of the proportion

π_{est}	Sample Size (n)
0.1	3458
0.2	6147
0.3	8068
0.4	9220
0.5	9604



Sample size determination for estimating the population variance

- Given that a random sample of size n is drawn from a normal distribution with mean μ and variance σ^2, the expression χ^2 follows a chi- square distribution with n-1 degrees of freedom, i.e.,

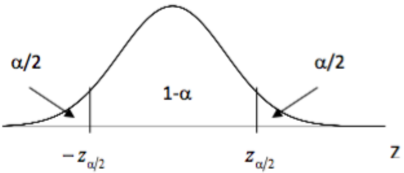
$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$$

- It can be shown that as $n \Rightarrow \infty$

  $\sqrt{2\chi^2}$ is approximately normally distributed with mean $\sqrt{2(n-1)}$ and variance 1.

- We will therefore determine the approximate sample size to estimate the population standard deviation, and hence the population variance, using percentiles of the standard normal distribution.

Sample size determination for estimating the population variance



$$P\left(-z_{\alpha/2} < \sqrt{2(n-1)s^2/\sigma^2} - \sqrt{2(n-1)} < z_{\alpha/2}\right) = 1-\alpha$$

$$\Rightarrow P\left(\left|\sqrt{2(n-1)s^2/\sigma^2} - \sqrt{2(n-1)}\right| < z_{\alpha/2}\right) = 1-\alpha$$

$$\Rightarrow P\left(\left|\sqrt{2(n-1)}\left(\frac{s-\sigma}{\sigma}\right)\right| < z_{\alpha/2}\right) = 1-\alpha$$

$$\Rightarrow P\left(\left|\frac{s-\sigma}{\sigma}\right| < \frac{z_{\alpha/2}}{\sqrt{2(n-1)}}\right) = 1-\alpha$$

Sample size determination for estimating the population variance

So that if the relative sampling error is to be less than E, then

$$E = \frac{z_{\alpha/2}}{\sqrt{2(n-1)}}$$

Therefore, the required sample size to estimate the population standard deviation is approximately

$$n = \frac{1}{2}\left(\frac{z_{\alpha/2}}{E}\right)^2 + 1$$

Example 3

- C batteries have a nominal diameter of 26.2mm.
- Producing batteries with this value will ensure they fit properly into all toys and appliances that require C batteries.
- A tolerance standard deviation of 0.395 mm of the diameter of the batteries is allowable.
- Suppose it is desired that the standard deviation of a sample of batteries is within 10% of the tolerance standard deviation with a confidence interval of 95%, how many batteries should be sampled.

$$E = 0.1, (1 - \alpha) = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow z_{\alpha/2} = z_{0.025}$$

$$n = \frac{1}{2}\left(\frac{z_{\alpha/2}}{E}\right)^2 + 1$$

We will now compute the sample size in R.

Example 3: R code and solution

```
> #sampling error
> E <- 0.1
> #97.5th percentile of the standard normal distribution
> z025 <- qnorm(.975)
> z025
[1] 1.959964
> #sample size
> n <- (0.5*(z025/E)^2) +1
> n
[1] 193.0729
```

Hence, the sample size that should be taken is at least 194.

Impact on sample size by varying components

Consider the sample size required to estimate the population standard deviation.

$$n = \frac{1}{2}\left(\frac{z_{\alpha/2}}{E}\right)^2 + 1$$

- If E are fixed, as the level of confidence increases
  - $z_{\alpha/2}$ increases; hence the sample size n increases.

- If the level of confidence is fixed, as the precision increases
  - E decreases; hence the sample size increases.

Varying the level of confidence

$$n = \frac{1}{2}\left(\frac{z_{\alpha/2}}{E}\right)^2 + 1$$

E is fixed, e.g., $E = 0.1$.

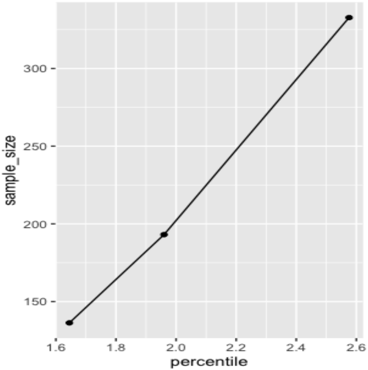
Level of confidence $(1 - \alpha)$ increases

$\Rightarrow$ Percentile, $z_{\alpha/2}$ increases

$\Rightarrow$ Sample size n increases

Table: Varying the level of confidence

$1-\alpha$	α	$z_{\alpha/2}$	n
90%	10%	1.644854	137
95%	5%	1.959964	194
99%	1%	2.575829	333



Varying the level of precision

$(1-\alpha)$ is fixed, e.g, $(1-\alpha) = 0.95$

Level of precision increases

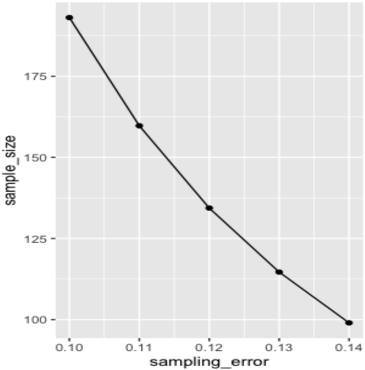
$\Rightarrow$ Sampling error, E decreases

$\Rightarrow$ Sample size n increases

$$n = \frac{1}{2}\left(\frac{z_{\alpha/2}}{E}\right)^2 + 1$$

Table: Varying the level of precision

Sampling Error (E)	Sample Size (n)
0.14	99
0.13	115
0.12	134
0.11	160
0.10	194



Factors affecting sample size

- So far we have discussed the determination of the appropriate sample size for estimation of the population parameters, mean, proportion, standard deviation (and hence variance), where in each cased we consider the following factors affecting the sample size.
  - Required level of precision
  - Required level of confidence
  - Preliminary estimate of variability in the case of estimating the population mean or population proportion.
- We will now work through an example to show how varying the cost impacts on the sample size.

Example 4

A market analyst in a particular city wants to estimate the proportion of shoppers who will buy a new type of dishwashing liquid that has been available in supermarkets over the past six months.

1. How large a sample should she take in order to estimate the proportion of shoppers who will purchase the new type of dishwashing liquid, with a maximum acceptable sampling error of 0.04 and with 95% confidence?
2. Suppose that a budget of $2,400 is available for this survey and the sample cost is $10 per shopper. Obtain the appropriate sample size that will meet the budget and obtain the sampling error for this sample size.
3. Compare the sampling costs and errors for the sample sizes obtained in Parts (1) and (2).

Example 4 - Part 1

1)
$E = 0.04, \pi_{est} = 0.5,$
$1 - \alpha = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow z_{\alpha/2} = z_{0.025}$

$$n = \left(\frac{z_{\alpha/2} \times \sqrt{\pi_{est}(1 - \pi_{est})}}{E} \right)^2$$

We will now compute the sample size in R.

Example 4: Part 1 - R code and solution

```
> #sampling error
> E <- 0.04
> #prelimary estimate of pi
> pi_est <- 0.5
> #97.5th percentile of the standard normal distribution
> z025 <- qnorm(.975)
> z025
[1] 1.959964
> #sample size
> n <- (z025*(sqrt(pi_est*(1-pi_est))/E)^2
> n
[1] 600.2279
```

Hence, the sample size that should be taken to ensure a sampling error of $E = 0.04$ is at least 601.

Example 4 - Part 2

Budget = \$2,400, Cost = \$10

Sample size that takes into account the budget and cost per shopper is:

n = budget/cost = 2400/10 = 240.

Given this sample size, we will now use R to compute the sampling error

$$E = z_{\alpha/2}\sqrt{\frac{\pi_{est}(1-\pi_{est})}{n}}.$$

Example 4: Part 2 - R code and solution

```
> #budget
> budget <- 2400
> #cost
> cost <- 10
> #sample size
> n <- budget/cost
> n
[1] 240
> #97.5th percentile of the standard normal distribution
> z025 <- qnorm(.975)
> z025
[1] 1.959964
> #sampling error
> E <- z025 * sqrt((pi_est*(1-pi_est))/n)
> E
[1] 0.06325757
```

Hence, a sample size of 240 shoppers will lead to a sampling error of approximately 0.06

Example 4: Part 3

Table: Comparison of sample sizes and corresponding sampling errors

Sample Size **n**	Sampling error **E**	**Cost**	
601	0.04	601 x 10 =	\$ 6,010
240	0.06	240 x 10 =	\$ 2,400

- Using the sample size as determined in Part (1), i.e., 601, increases the cost by (\$6010 - \$2400)/\$2400 = 150%.
- The sampling error when operating within the budget, i.e., \$2400 increases by about (0.06 - 0.04)/0.04 = 50% over that allowed in Part (1).
- Working within the budget of \$2,400 will result in a lower quality estimate, while trying to maintain the sampling error of E = 0.04, increases the cost by 150%.
- Hence for the marketing research company, a balance will have to be struck between cost and the quality of the estimate of proportion.

Summary

Sample size determination for estimating the population mean.

- For an *infinite* population

$$n = \left(\frac{z_{\alpha/2} \times \sigma_{est}}{E} \right)^2$$

- For a finite population

$$n = \frac{N(z_{\alpha/2} \times \sigma_{est})^2}{(N-1)E^2 + (z_{\alpha/2} \times \sigma_{est})^2}$$

Summary

Sample size determination for estimating the population proportion

- For an *infinite* population

$$n = \left(\frac{z_{\alpha/2} \times \sqrt{\pi_{est}(1 - \pi_{est})}}{E} \right)^2$$

- For a finite population

$$n = \frac{n_0}{1 + \frac{n_0 - 1}{N}}$$

$$n_0 = \left(\frac{z_{\alpha/2} \times \sqrt{\pi_{est}(1 - \pi_{est})}}{E} \right)^2$$

- If there is no preliminary estimate of proportion, then $\pi_{est} = 0.5$ can be used.

Summary

The required sample size to estimate the population standard deviation and hence the population variance is approximately

$$n = \frac{1}{2}\left(\frac{z_{\alpha/2}}{E}\right)^2 + 1.$$

Summary

Factors affecting sample size

- Required level of precision
- Required level of confidence
- Preliminary estimate of variability (population standard deviation or population proportion)
- Cost involved