# SAMPLE SIZE\*

# Mary Teegarden

This work is produced by OpenStax-CNX and licensed under the Creative Commons Attribution License  $2.0^{\dagger}$ 

#### Abstract

Calculations for determining the required sample sized when calculation a confidence interval for the population mean or population proportion.

## 1 Determining Sample Size Required to Estimate $\mu$

Prior to creating a confidence interval a sample must be taken. Often the number of data values needed in a sample to obtain a particular level of confidence within a given error needs to be determined prior to taking the sample. If the sample is too small the result may not be useful and if the sample is too big both time and money are wasted in the sampling.

Let's begin by looking at the equation for the Error Bound.

 $\mathbf{EBM} = Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ 

To affect the size of the error, what can be changed in the formula?

Only  $Z_{\frac{\alpha}{2}}$ , by changing the level of confidence, or n by changing the sample size affects the error. The standard deviation is a given which one can not change. (Note if n > 30 then the sample standard deviation can be used to approximate the population standard deviation.)

### Example 1

Given the following data: n = 64,  $\overline{x} = 36$ , and  $\sigma = 3$ , the EBM for a 80%, 90%, 95% and 99% confidence interval are:

80%: EBM = 
$$Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = 1.28 \frac{3}{\sqrt{64}} = 0.48$$
  
90%: EBM =  $Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = 1.645 \frac{3}{\sqrt{64}} = 0.616875$   
95%: EBM =  $Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = 1.96 \frac{3}{\sqrt{64}} = 0.735$   
99%: EBM =  $Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = 2.58 \frac{3}{\sqrt{64}} = 0.9675$ 

Note that as the confidence increases, so also does the EBM. To ensure that the error bound is small, the confidence must be decreased. Hence changing the confidence to lower the error is not a practical solution.

### Example 2

What happens as the sample size is increased? Calculate the EBM for a 90% confidence interval for n = 25, 64, and 100.

n = 25: EBM = 
$$Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = 1.645 \frac{3}{\sqrt{25}} = 0987$$
  
n = 64: EBM =  $Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = 1.645 \frac{3}{\sqrt{64}} = 0.616875$   
n = 100: EBM =  $Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = 1.645 \frac{3}{\sqrt{100}} = 0.4935$ 

<sup>\*</sup>Version 1.2: Jul 14, 2009 9:58 pm -0500

<sup>&</sup>lt;sup>†</sup>http://creativecommons.org/licenses/by/2.0/

As the sample size increases, the EBM decreases. The question now becomes how large a sample is needed for a particular error?

Begin by solving the equation for the EBM in terms of n.

$$\text{EBM} = Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \Rightarrow n = \left(\frac{Z_{\frac{\alpha}{2}}\sigma}{\text{EBM}}\right)^2 \tag{1}$$

- Where Z = the critical z score based on the desired confidence level
- EBM = desired margin of error
- $\sigma =$ population standard deviation

Often the population standard deviation is unknown; hence the sample standard deviation from a previous sample of size greater than 30 may be used as an approximation to  $\sigma$ .

The value found by using the formula for sample size is generally not a whole number. However the sample size must be a whole number, so ALWAYS ROUND UP to the next larger whole number.

### Example 3

Suppose for the given information, we want to be 90% confident with an error of only  $\pm 0.5$ , how large should n be?

Now substitute the given values:

$$n = \left(\frac{Z_{\frac{\alpha}{2}}\sigma}{\text{EBM}}\right)^2 = \left(\frac{1.645\,(3)}{0.5}\right)^2 = 9.87^2 = 97.4169\tag{2}$$

Since you can not sample less than a whole this value n = 98.

### Example 4

How large a sample would you need if you wanted to be 95% confident with an error of  $\pm 0.25$ ?

$$n = \left(\frac{Z_{\frac{\alpha}{2}}\sigma}{\text{EBM}}\right)^2 = \left(\frac{1.96(3)}{0.25}\right)^2 = 23.52^2 \approx 554.19 \Rightarrow n = 554 \tag{3}$$

### Exercise 1

(Solution on p. 4.)

Suppose the scores on a statistics final are normally distributed with a standard deviation of 10 points. You have been asked to construct a 95% confidence interval with an error of no more than 2 points.

### Exercise 2

(Solution on p. 4.)

Suppose you want to be 98% confident with an error of no more than 1.5 points, how large must your sample be?

### 2 Determining Sample Size Required to Estimate p.

Again, let's look at the equation for the Error Bound.

$$\text{EBP} = Z_{\frac{\alpha}{2}} \sqrt{\frac{p'q'}{n}}$$

- where  $\mathbf{p}' = \frac{x}{n}$  is the point estimate for the true population proportion
- $\mathbf{x} = \mathbf{the} \ \mathbf{number} \ \mathbf{of} \ \mathbf{successes}$
- n =the sample size
- q' = 1 p'

OpenStax-CNX module: m17343

Solving for n, we obtain:  $n = \left(\frac{Z_{\frac{\alpha}{2}}}{\text{EBP}}\right)^2 p'q'$ 

### Example 5

Suppose that a previous study claimed that only 25% of people recycle on a regular basis. Determine the sample size needed to create a 95% confidence interval for the true population proportion with an error of only +/-3%.

$$\begin{aligned} \mathbf{p}' &= 0.25 \implies \mathbf{q}' = 1 - \mathbf{p}' = 0.75\\ n &= \left(\frac{Z_{\frac{\alpha}{2}}}{\text{EBP}}\right)^2 p' q' = \left(\frac{1.96}{0.03}\right)^2 (0.25) (0.75) = 800.33\\ \text{AWAYS ROUND UP, hence } \mathbf{n} = 801. \end{aligned}$$

# If there is no previous sample, let p' = 0.5 and q' = 0.5 as this gives the largest value for n.

### Example 6

What would n be for a 95% confidence interval with an error of only +/-3% with no sample data?

$$n = \left(\frac{Z_{\frac{\alpha}{2}}}{\text{EBP}}\right)^2 p' q' = \left(\frac{1.96}{0.03}\right)^2 (0.5) (0.5) \approx 1067.1$$
(4)

AWAYS ROUND UP, hence n = 1068

Note that not having a previous sample greatly increases the number of data values needed in a sample. Often a pilot study is done to generate an approximation for p.

### Exercise 3

### (Solution on p. 4.)

(Solution on p. 4.)

(Solution on p. 4.)

The Mesa College mathematics department has noticed that a number of students place in a nontransfer level course and only need a 6 week refresher rather than an entire semester long course. If it is thought that about 10% of the students fall in this category, how many must the department survey is they wish to be 95% certain that the true population proportion is within +/-5%?

### Exercise 4

Suppose the math department has no previous information. How many students should be surveyed?

### Exercise 5

Suppose Cardmart wish to know what proportion of men buy their wife a Mother's Day Card. How many people must be sampled is they wish to be 95% certain that the proportion is within 2%? Suppose that a previous sample of 500 men reported that 421 of them bought their wife a Mother's Day Card.

### Exercise 6

# (Solution on p. 4.)

Suppose there is no previous sample. How many men will need to be surveyed?

# Solutions to Exercises in this Module

### Solution to Exercise (p. 2)

 $Z_{0.025} = 1.645$ EMB = 2  $\sigma = 10$  $n = \left(\frac{Z_{\frac{\alpha}{2}}\sigma}{\text{EBM}}\right)^2 = \left(\frac{1.645(10)}{2}\right)^2 = (8.225) 2 = 67.65$  rounded up => n = 68.

Hence, a sample of size 68 must be taken to create a 95% confidence interval with an error of no more than two points.

Solution to Exercise (p. 2)  $Z_{0.01} = 2.33$ EMB = 1.5

 $\sigma = 10$ 

$$n = \left(\frac{Z_{\frac{\alpha}{2}}\sigma}{\text{EBM}}\right)^2 = \left(\frac{2.33\,(10)}{1.5}\right)^2 = (15.533)\,2 = 241.28 \text{ rounded up} => n = 242.$$
(5)

Hence, a sample of size 242 must be taken to create a 98% confidence interval with an error of no more than 1.5 points.

Solution to Exercise (p. 3)

$$n = \left(\frac{Z_{\frac{\alpha}{2}}}{\text{EBP}}\right)^2 p' q' = \left(\frac{1.96}{0.05}\right)^2 (0.1) (0.9) \approx 138.3$$
(6)

AWAYS ROUND UP, hence n = 139Solution to Exercise (p. 3)

$$n = \left(\frac{Z_{\frac{\alpha}{2}}}{\text{EBP}}\right)^2 p' q' = \left(\frac{1.96}{0.05}\right)^2 (0.5) (0.5) \approx 384.2$$
(7)

AWAYS ROUND UP, hence n = 385Solution to Exercise (p. 3)

$$\mathbf{p}' = \frac{421}{500} = 0.842\tag{8}$$

, hence

$$q' = 1 - p' = 1 - 0.842 = 0.158$$
 (9)

 $E\,=\,0.02$ 

$$n = \left(\frac{Z_{\frac{\alpha}{2}}}{\text{EBP}}\right)^2 p' q' = \left(\frac{1.645}{0.02}\right)^2 (0.842) (0.158) \approx 899.997$$
(10)

AWAYS ROUND UP, n = 900. Hence 900 people need to be surveyed to ensure a 95% confidence interval with an error of at most 2%.

Solution to Exercise (p. 3)

$$n = \left(\frac{Z_{\frac{\alpha}{2}}}{\text{EBP}}\right)^2 p' q' = \left(\frac{1.645}{0.05}\right)^2 (0.5) (0.5) \approx 1691.27 \tag{11}$$

AWAYS ROUND UP, n = 1692. Hence 1692 people need to be surveyed to ensure a 95% confidence interval with an error of at most 2%.