Sample Size

Suppose we wish to construct a confidence interval for a mean $\mu$ with level of confidence $r$. However we wish to have a reasonably small margin of error $e$ in order to obtain a useful interval. What random sample size $n$ will guarantee a margin of error of no more than $e$ regardless of the measurements we are studying? If we want the margin error to be no more than $e$, then we must set

$$\frac{z_r S}{\sqrt{n}} \leq e$$

Then we solve for $n$ to obtain $n \geq \left(\frac{z_r S}{e}\right)^2$. But when $S$ is unknown in advance, we cannot use $S$ in this formula because we are trying to determine the required sample size before taking the sample. In that case, we should use an upper bound $U$ for $S$.

**Upper Bound $U$ for $S$**

If we know the range $[c, d]$ of the possible measurements under study, then we can find an upper bound for $S$. Among all measurements with range $[c, d]$, the maximum possible standard deviation is $\frac{d-c}{2}$. Thus for a population measurement with range $[c, d]$, an upper bound for $S$ is given by $U = \frac{d-c}{2}$.

For example, suppose a minimum GPA of 2.8 is required of applicants into the Nursing Program. Then $2.8 \leq X \leq 4.0$, where $X$ is the GPA of all applicants to be considered. Then $U = (4.0 - 2.8)/2 = 0.6$, and so $S \leq 0.6$.

**Choosing the Sample Size**

The required sample size needed to obtain a margin of error of no more than $e$ with level of confidence $r$ is now given by

$$n \geq \left(\frac{z_r U}{e}\right)^2 \text{ where } U = \frac{d-c}{2}.$$ 

**Note:** $n$ is always rounded up to the nearest integer.

In reality, the above sample size formula is only good for very large populations. But we can improve the formula if we actually know the entire population size $N$. In this case, we use

$$n \geq \frac{N\left(\frac{z_r U}{e}\right)^2}{N-1+\left(\frac{z_r U}{e}\right)^2} \text{ where } U = \frac{d-c}{2}.$$
Example 1. Acceptance into the Business College requires a minimum GPA of 2.5. We want to find a 95% confidence interval for the mean GPA of all applicants that has no more than a 0.05 margin of error.

(a) Find the required sample size from an extremely large population of applicants.
(b) Find the required sample size from a pool of only 900 applicants.

Solution. First, an upper bound for $S$ is $U = (d - c) / 2 = (4 - 2.5) / 2 = 0.75$.

(a) Now using a $z$–score of 1.96 for 95% confidence and a margin of error of $e = 0.05$, the required sample size is $n \geq \left( \frac{1.96 \times 0.75}{0.05} \right)^2 = 29.4^2 = 864.36$, or 865 people.

(b) With a population size of $N = 900$, the required sample size is

$$\frac{900 \left( \frac{1.96 \times 0.75}{0.05} \right)^2}{899 + \left( \frac{1.96 \times 0.75}{0.05} \right)^2} = \frac{900 \times 864.36}{899 + 864.36} \approx 441.16.$$ 

So we require a random sample of size $n = 442$ from the 900 applicants.

Practice Exercises

1. Suppose we want a maximum margin of error of 20 when finding a 99% confidence interval for the average GRE of Graduate School applicants, where a minimum score of 1200 out of a possible 1600 is required. Find the necessary random sample size for such a confidence interval in the following cases:

(a) Among a huge pool of applicants nationwide.
(b) Among a select pool of 500 applicants.

2. We wish to find a 98% confidence interval for the average number of years needed to graduate for all 1990’s alumni. Excluding the rare extreme cases, all graduates took from 3 to 8 years to graduate.

(a) If we wanted to estimate the average number of years needed with a maximum margin of error of 0.25 with 98% confidence, what sample size would be needed?

(b) What sample size of the 350 chemistry majors would be needed to estimate the average number of years they needed with a maximum margin of error of 0.25 with 98% confidence?
Sample Size for a Proportion

As with confidence intervals for the mean, we may like to know in advance what sample size would provide a certain maximum margin of error $e$ with a certain level of confidence $r$. For a proportion, all measurements are either 0 (for No) or 1 (for Yes). So now $U = (1 - 0)/2 = 0.5$. Hence, the required sample size $n$ is given by

$$n \geq \begin{cases} \left( \frac{z_r \times 0.5}{e} \right)^2 & \text{for large populations} \\ N \times \left( \frac{z_r \times 0.5}{e} \right)^2 & \text{for a population of size } N \\ (N - 1) + \left( \frac{z_r \times 0.5}{e} \right)^2 & \end{cases}$$

where $z_r$ is the appropriate $z$-score depending on the level of confidence. Note: We always round up to the nearest integer.

Example 2. What sample size will guarantee a maximum margin of error of 0.035 for any 99% confidence interval of a proportion? What sample size would guarantee the result from a population of size 1200?

Solution. For a large population, the required sample size must satisfy

$$n \geq \left( \frac{z_{\alpha/2} \times 0.5}{e} \right)^2 = \left( \frac{2.576 \times 0.5}{0.035} \right)^2 = 1354.24;$$

thus, $n$ must be at least 1355. From a population of size 1200, the sample size must satisfy

$$n \geq \frac{N \times \left( \frac{z_{\alpha/2} \times 0.5}{e} \right)^2}{(N - 1) + \left( \frac{z_{\alpha/2} \times 0.5}{e} \right)^2} = \frac{1200 \times 1354.24}{1199 + 1354.24} = 636.48$$

thus, $n$ must be at least 637.

More Practice Exercises

3. (a) In a nationwide survey, if you wanted to estimate a proportion within 0.025 with 95% confidence, then how many people must be surveyed?

(b) Suppose now that you are only surveying a population of 1200 people. If you wanted to estimate the same proportion $p$ within 0.25 with 98% confidence, then how many of the 1200 would you need to survey?
Answers

1. (a) Here \( U = \frac{1600 - 1200}{2} = 200 \). Need \( n \geq \left( \frac{2.576 \times 200}{20} \right)^2 = (25.76)^2 = 663.5776 \).
   So sample 664 people.

   (b) \( n \geq \frac{500 \left( \frac{2.576 \times 200}{20} \right)^2}{499 + \left( \frac{2.576 \times 200}{20} \right)^2} = \frac{500 \times (25.76)^2}{499 + (25.76)^2} \approx 285.39 \); so sample 286 out the 500.

2. (a) Here, \( U = \frac{8 - 3}{2} = 2.5 \). So we need \( n \geq \left( \frac{2.326 \times 2.5}{0.25} \right)^2 = (23.26)^2 \approx 541.0276 \).
   So sample 542 people.

   (b) \( n \geq \frac{350 \left( \frac{2.326 \times 2.5}{0.25} \right)^2}{349 + \left( \frac{2.326 \times 2.5}{0.25} \right)^2} = \frac{350 \times (23.26)^2}{349 + (23.26)^2} = 212.757 \); so sample 213 of the 350.

3. (a) \( n \geq \left( \frac{1.96 \times 0.5}{0.025} \right)^2 = 39.2^2 = 1536.64 \); so sample 1537.

   (b) \( n \geq \frac{1200 \left( \frac{1.96 \times 0.5}{0.025} \right)^2}{1199 + \left( \frac{1.96 \times 0.5}{0.025} \right)^2} = \frac{1200 \times 39.2^2}{1199 + 39.2^2} \approx 674.0536 \); so sample 675 of the 1200.