

Calculating a Large Sample Confidence Interval for a Population Mean

Example. Recall the results of EPA tests to determine mileage ratings for new car models. The data we examined represented the results of 100 tests on a certain new car model. That data is contained in the EPAGAS file on the CD that came with our text. Suppose we are interested in estimating the mean MPG rating for that new car model, and we choose to proceed by basing our estimate on a random sample of the 100 measurements in the data file.

- We use MINITAB to select a random sample of size $n = 30$ from the data set EPAGAS. (See page 166 for instructions on generating a random sample in MINITAB.)
- We find the mean \bar{x} and standard deviation s of the 30 MPG ratings we obtained in part a.

Variable	N	Mean	StDev	SE Mean
Sample	30	37.003	2.074	0.379

- We use the information in part b to form a large sample 99% confidence interval for the true mean MPG rating for the particular car model.
- To calculate the large sample 99% confidence interval for μ we use the following form:

$$\begin{aligned}\bar{x} \pm z_{\alpha/2} \left(\frac{s}{\sqrt{n}} \right) &= \bar{x} \pm z_{0.005} \left(\frac{s}{\sqrt{n}} \right) = \bar{x} \pm 2.575 \left(\frac{s}{\sqrt{n}} \right) \\ &\approx 37.003 \pm 2.575 \left(\frac{2.074}{\sqrt{30}} \right) \approx 37.003 \pm 0.975\end{aligned}$$

So, our 99% confidence interval is (36.028, 37.978)

- Now, we give a practical interpretation of that confidence interval.

Calculating a Small Sample Confidence Interval for a Population Mean

Example: We repeat the example above but with a small sample of size $n = 10$.

- In this case the relevant MINITAB output is

Variable	N	Mean	StDev	SE Mean
Sample 2	10	37.400	1.573	0.497

- We use the information in part a to form a small sample 99% confidence interval for the true mean MPG rating for the particular car model.
- To calculate a small sample 99% confidence interval for μ we use the following form:

$$\bar{x} \pm t_{\alpha/2} \left(\frac{s}{\sqrt{n}} \right) = 37.4 \pm t_{.005} \left(\frac{1.573}{\sqrt{10}} \right) = 37.4 \pm 3.250 \left(\frac{1.573}{\sqrt{10}} \right) \text{ where } t_{.005} \text{ is based on } (n - 1) \text{ degrees of freedom (df).}$$

So, our small sample 99% confidence interval is (35.783, 39.017)

Variable	N	Mean	StDev	SE Mean	99.0 % CI
Sample 2	10	37.400	1.573	0.497	(35.783, 39.017)

How to Determine Sample Size

Looking back at our calculation of a 99% confidence interval for the true mean MPG rating for a particular car model, we can consider the effect the sample size n has on the width of the confidence interval. We note that half the width of the confidence interval is given by

$$\bar{x} \pm z_{\alpha/2} \left(\frac{\sigma}{\sqrt{n}} \right) \approx \bar{x} \pm z_{0.005} \left(\frac{s}{\sqrt{n}} \right) \approx 37.003 \pm 2.575 \left(\frac{2.074}{\sqrt{n}} \right)$$

We can see that if the sample size n increases, then the width of our confidence interval will shrink. Our calculation of the large sample 99% confidence interval yielded an interval of width $2(0.975)=1.95$. Suppose we had wanted to obtain an estimate that was within 0.75 of the true mean. That is, we had wanted a 99% confidence interval of width 1.5. How large a sample would have been required?

Assuming that the true population σ is unknown, and we use a sample's standard deviation s as an estimate for σ , we can find the sample size n required for half the width of the 99% confidence interval to be 0.75 by solving the following equation for n .

$$2.575 \left(\frac{2.074}{\sqrt{n}} \right) = 0.75$$

Solving the equation we find $n \approx 51$.

We express the reliability associated with a confidence interval for the population mean μ by specifying the **sampling error**, denoted by SE, within which we want to estimate μ with $100(1-\alpha)\%$ confidence. In the above example, $SE = 0.75$.

To estimate μ with a sampling error SE with $100(1 - \alpha)\%$ confidence, the required sample size is found by solving for n in the following equation.

$$z_{\alpha/2} \left(\frac{\sigma}{\sqrt{n}} \right) = SE$$

The resulting solution for n is

$$n = \frac{(z_{\alpha/2})^2 \sigma^2}{(SE)^2}.$$

Since the value of σ is usually not known, we can estimate its value by the standard deviation s of a sample. As an alternative, we can estimate the range R of observations in the population and use the conservative estimate $\sigma \approx R/4$.

Example: We wish to estimate the mean of a normal population with a 95% confidence interval and we know from prior information that the population variance is 1.

- Investigate the effect of sample size on the width of the confidence interval, by calculating the width of the confidence interval for $n = 16, 25, 49, 100,$ and 400 .
- Plot the width as a function of the sample size n . Connect the points with a smooth curve and note how the width decreases as n increases.
- What minimum sample size will determine a confidence interval of width 0.50?